SIMULATION OF SIMILARITY AND PREVIOUS KNOWLEDGE GESTALT RULES WITH COUPLED NEURAL OSCILLATORS

E. Magosso, G.E. La Cara, C. Cuppini, M. Ursino

Department of Electronics, Computer Science and Systems, University of Bologna, Bologna, Italy

emagosso@deis.unibo.it

Abstract: Object recognition requires the solution of the binding and segmentation problems, i.e., grouping different features of an object to achieve a coherent representation. Synchronization of neural activity in the γ-band has often been proposed as a putative mechanism to solve these problems. In the present work, a network of Wilson-Cowan oscillators is used to segment simultaneous objects, and recover an object from a partial or corrupted information, by implementing two gestalt rules: similarity and previous knowledge. The network consists of H different areas, each devoted to representation of a particular feature of the object, according to a topological organization. The similarity law is realized via lateral intra-area connections, arranged as a "Mexican-hat". The previous knowledge is realized via inter-area connections, which link properties belonging to a previously memorized object. A global inhibitor allows segmentation of several objects. Simulation results, performed using three simultaneous input objects, show that the network is able to detect an object even in difficult conditions (i.e., when some features are absent or corrupted). Objects with excessive corruption are correctly rejected. The network exhibits a good compromise between sensitivity (capacity to detect true positives) and specificity (capacity to reject false positives).

Introduction

Execution of many cognitive functions requires that different features of perception are grouped together (binding problem) to permit the recognition of natural scenes and the achievement of coherent object representation. Of course, features of different objects, simultaneously present in the same scene, must be maintained distinct and independently processed (segmentation problem).

In general, the neural system utilizes a limited number of features to classify and recognize perceived objects [1]. An early hypothesis is that the presence of an object is signaled by a specialized neuron, which would process individual features via a feed-forward and hierarchically structured process, and would encode increasingly complex relationships [2]. According to this idea, the simultaneous presence of two objects in the same scene is signaled by activation of two distinct specialized neurons. This mechanism, however, exhibits several drawbacks, and is generally rejected in the neurophysiological literature today. First, considering all possible combinations of features lead to a combinatorial explosion of possibilities, hence to an excessive number of individual neurons. Furthermore, with this mechanism, it is difficult to incorporate new knowledge and to deal with entirely novel objects [3].

The previous limitations may be overcome by the so-called "assembly coding". In this mechanism the presence of an object is not signaled in the nervous system by individual units, but rather by the concurrent activity of many neurons simultaneously excited, each signaling a single feature, located in proximal or distal cortical areas and linked via functional reciprocal connections [1]. Two fundamental problems, however, arise in this context. How segmentation can be achieved in assembly coding, avoiding that features of two distinct objects are erroneously grouped together? What are the fundamental cues exploited to generate this segmentation, i.e., to group simultaneously active neurons into distinct objects?

One possible solution to the first problem assumes that different features of an object are grouped together via synchronization of neural firing [4-8]. Cortical neurons, in fact, are often engaged in synchronous activity in the γ-frequency range (40-60 Hz) [6,9]. This synchronization may be distributed both within the same cortical area and among distant areas, and is not locked to external stimuli, i.e., it depends on internal connections among neurons (i.e., on an internal representation of objects). According to synchronization hypothesis, neurons that fire in phase would signal attributes of the same object, while neurons firing out of phase would signal attributes in different objects. This hypothesis, although still debated in the literature and not universally accepted [10] has received important experimental support recently from a variety of studies [11-13].

As to the second question, fundamental ideas on the sensory cues that may be employed for segmentation are provided by the Gestalt Psychology [14]. Gestalt grouping rules can be of low-level, reflecting basic properties of sensory inputs (such as closeness, smoothness, boundaries and common fate), or high-level, reflecting more complex features extracted from a pre-processing system (such as similarity and previous knowledge).

The role of neural synchronization in binding and segmentation can be critically analyzed using mathematical models and computer simulation techniques. Indeed, many models of oscillating neural

ISSN: 1727-1983 © 2005 IFMBE

networks, with a different level of complexity and of physiological reliability have been proposed in past years, with encouraging results. In these models, the rules used for segmentation are generally inscribed into the synaptic connections linking oscillators. However, most of these studies are focused on low-levels Gestalt cues, such as proximity, smoothness and common fate to segment a visual scene at an early processing visual stage [15-17], whereas just a few attempts to use high-levels rules to classify more complex objects at a higher mental level have been performed [4,5,18].

The objective of the present work is to develop a neural network of oscillating units, focusing attention especially on segmentation using high-level cues with possible emphasis on higher cortical functions. The essential concept of our model is that classification and representation of high-level objects may be realized starting from a partial or incomplete sensory information, by grouping together a limited set of fundamental features or attributes. We assume that these basic features are extracted from sensory perception at an earlier processing stage, and are arranged in a topologically ordered fashion at some areas of the cortex. Features are then linked together (binding) and separated (segmentation) by synchronization in the γ-range using the similarity and previous knowledge Gestalt rules, in order to arrive at high-level (semantic) object representation.

To illustrate the main ideas of our model, we propose a simple implementation, in which complexity is intentionally maintained at a minimum level. The objective is to show how this network may work, its virtues and robustness. More complex and physiologically founded networks may be naturally built in subsequent works.

Materials and Methods

We assume that the model is composed of N oscillating neural groups, subdivided into H distinct cortical areas. Each neural group may be silent, if it does not receive enough excitation, or may oscillate in the γ-frequency band, if excited by a sufficient input. A single oscillator in the network is described by means of Wilson-Cowan equations, with coupling terms chosen to allow fast and efficient synchronization and a global inhibitor to allow desynchronization among properties in different objects [17].

A schematic representation of the main topological aspects of the network, is presented in Fig. 1.

Each area is devoted to the representation of a specific attribute or feature of the object (for instance color, orientation, geometrical form in case of visual stimuli, tone in case of auditory stimuli, body position in case of somatosensory stimuli, etc…). Hence, one object is represented as the collection of H features (one feature per each area). We assume that each attribute is not immediately present in the sensory input, but has been extracted from a previous processing stage in the neocortex.

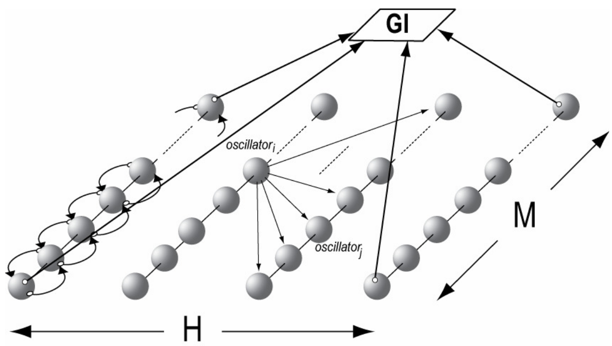

Figure 1: Schematic diagrams describing the model structure. The model is composed of *H* different cortical areas each represented as a mono-dimensional chain of *M* Wilson-Cowan oscillators. Each oscillator receives coupling terms both from oscillators in the same area (lateral intra-area connections), and from oscillators in different areas (inter-area synapses). Moreover each unit of the network receives an inhibitory signal from the global inhibitor (GI).

Neural groups within each area represent the value of that particular attribute according to a topological organization. This means that two proximal neural groups in the area signal the presence of two similar values, while distant groups signal the presence of different values. This topological organization is very frequent in the neocortex to represent sensory modalities (let us consider, for instance, the orientation map or the color map in the visual cortex, the tonotopic map in the auditory cortex, etc…).

Neural groups within the same area are connected via lateral excitatory and inhibitory synapses (say *L* in the subsequent text). These lateral connections are organized according to a classical "Mexican hat" disposition. This means that a neuron excites (and is excited by) its proximal neurons in the area, whereas it inhibits (and is inhibited by) more distal neurons. As it is well known, excitatory neurons in the cortex may inhibit proximal neurons in the same area via inhibitory interneurons. Hence, all negative synapses within each area are realized via a bisynaptic connections, from excitatory units to inhibitory units, and then from the latter to other excitatory units.

Two neural groups belonging to different areas may be connected via symmetrical excitatory synapses (say *W* in the subsequent text). These reflect the existence of long range functional connections among different cortical areas. These synapses are normally equal to zero, but may assume a positive value when the two neural groups have been simultaneously active in the past during the learning phase. Hence, these synapses store a "previous knowledge" on whether different attributes occurred together in the past during the presentation of objects.

Lateral intra-area connections implement a similarity criterion, i.e., neural groups which signal a similar value for the attribute tend to be simultaneously active. Inter-area synapses implement a previous knowledge criterion, i.e., attributes which were collected together in the past tend to be grouped again in future experience.

Finally, in self-organizing networks the input to a neuron is generally computed as the scalar product between a sensory vector and the vector of synapses entering the neuron [19]. In the present study, for the sake of simplicity, the input to each neuron is described as a scalar quantity, ranging between 0 and 1, which reflects the similarity of the input with the value signaled by the given neuron.

Results

Simulations have been performed assuming that three objects are simultaneously given as input to the network. Each object is represented by four exact attributes during the "storage" phase. These attributes are stored in the matrix W of the inter-area synapses, reflecting previous knowledge. The presence of lateral connections, L, which cannot be modified by experience, produces an "activation bubble" within each area, i.e., not only the stimulated neurons oscillate, but also neurons in the same area signaling similar properties. The width of the excitation bubble, hence the degree of specificity depends on a balance between lateral excitation and lateral inhibition. In this condition, synapses W in the model, which incorporate previous knowledge, not only ensure a rapid synchronization between properties of the same object, but also allow restoration of lacking information. In order to underline this aspect, during the "retrivial" phase we will assume that objects can be presented in incomplete form (i.e., lacking some attributes) and/or in a corrupted form (i.e., some attributes may be a little different from the exact ones). Aim of the network is to recognize previously learned objects, despite a certain degree of incompleteness and/or corruption.

The first simulations (figure 2) have been performed assuming the absence of one property in each object. The other three properties are stimulated with an input $I_i = 0.8$. The figure shows network activity in all neural groups at different snapshots during the simulation. The network recovers the lacking property in each object; in other words, the object can be completely reconstructed, re-creating the property which is not given as input.

Figure 3 shows the same simulation, assuming the absence of two properties in objects 2 and 3 (i.e., only 2 properties over 4 are given as to these objects). With the basal value for the synapses W, the information is insufficient to recuperate the entire object.

Just two properties may be sufficient to recover an entire object from previous knowledge, if we assume a stronger value for the synapses W. The simulation results are not shown for the sake of brevity, since they are almost indistinguishable from those presented in Fig. 2.

In conclusion, the previous simulations illustrate the possibility to reconstruct an entire object from previous knowledge starting from partial information, still satisfying the binding and segmentation problem.

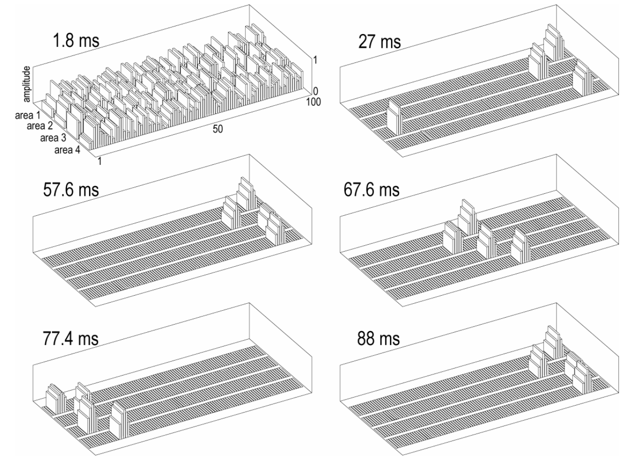

Figure 2: Network activity at different snapshots during the numerical simulation. Each pixel represents an oscillator. The emerging height is proportional to the corresponding oscillator's activity. In the simulation, three objects are present in the sensory input, each lacking of one attribute. After a brief transient from an initial random condition, the three objects are perfectly reconstructed by the network, recovering the fourth lacking property. Separation among the three objects is achieved via synchronisation of neurons responding to the same object and desynchronisation of neurons coding for different objects.

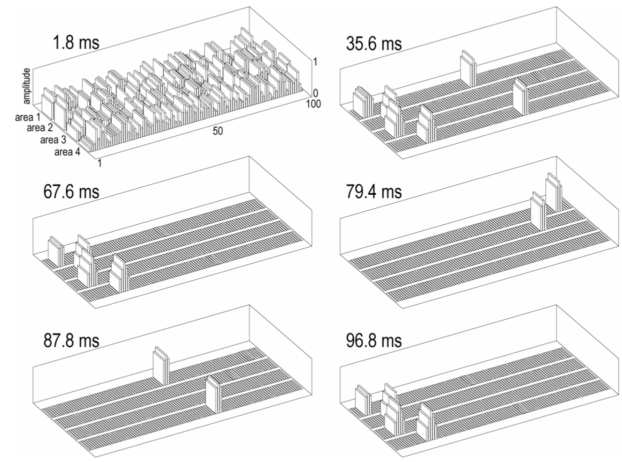

Figure 3: Network activity at different snapshots during the simulation. Simulation is similar to that of Fig.2. However, in this case, we assumed the absence of one property in one object, and the absence of two properties in the other two objects. The network is not able to recover the two lacking properties, hence only one object is correctly reconstructed.

Reconstruction from partial information depends on information stored in the synaptic matrix W. The higher the values of the trained synapses the smaller the number of properties necessary to recover an incomplete object.

Further simulations were performed assuming that some attributes are corrupted from the "exact" value. These simulations are summarized in Table 1. This table shows the percentage of success in 10 different trials (with random initial values for the network) and the settling time, i.e., the time required for achieving a synchronization.

In the first simulation (Table 1, first column) we assumed that the network receives two correct properties for each object. We remind that, according to Fig. 3, two properties are insufficient to recover the entire object. However, we now assume that object 1 also receives a property that is shifted by just 1 position from one of the lacking properties. Moreover, we assume that also objects 2 and 3 receive a "corrupted" property, which differ by just 1 position from the exact one. Thanks to lateral connections, the existence of this "similar" property is sufficient to evoke the overall object, including the fourth lacking property. Table 1 shows that the percentage of success is 90%. The time required for achieving a synchronization is short (average 30-40 ms).

Table 1: Percentage of success in 10 trials for three simulated conditions, and average time required to achieve synchronisation. In the three simulated conditions (A, B and C), all the three objects are present in the sensory input. Each object receives two exact properties and a third corrupted property which may differ by one position (configuration *i*) or by two positions (configuration *ii*) from the exact value (the fourth property is lacking). In condition A, configuration *i* holds for all the three objects. In condition B, configuration *i* holds for two objects, whereas configuration *ii* holds for one object. In condition C, configuration *i* holds for one object, and configuration *ii* holds for two objects. The desired behaviour of the network (success event) is the recognition of the objects belonging to configuration *i*.

condition A	**condition B**	**condition C**
recognition of three objects	recognition of two objects	recognition of one object
9/10	10/10	9/10
36.1 ms	38.5 ms	49.11 ms

However, if one property is shifted by 2 from the original one (Table 1, second and third columns), the object cannot be reconstructed. Nevertheless, the remaining objects (with a property corrupted by just one position) are correctly reconstructed. Hence, the network works well to reconstruct objects with a low degree of corruption, avoiding reconstruction of objects with excessive corruption.

Discussion

Objects are defined as a collection of different features, which must be grouped together to achieve a correct object reconstruction, but must be taken apart from features of different objects to avoid confusion. Moreover, these features are processed in distinct areas of the brain, and are generally reproduced via a topologically ordered organization. The problem still remains open on how the brain can integrate this sparse and highly distributed information to achieve a coherent and cohesive perception of the external world. Several authors in past years have linked fast oscillatory activity to learning and memory, especially in the perception of previously recognized objects [11,20,21].

Aim of this work is to propose a simple model for high-level object representation, which exploits two fundamental Gestalt rules: previous knowledge and similarity, together with synchronization among oscillatory neural populations. Previous knowledge is incorporated into the model in the synapses linking properties in one area to properties in another area. The similarity principle ensues from the presence of lateral (excitatory and inhibitory) synapses within the same area, which are arranged according to the classical "Mexican hat". The consequence of this specific disposition of synapses is that excitation of a neural group causes the occurrence of an excitation bubble, i.e., activation of one feature is always associated with the activation of similar features in the same area. As a consequence, similarity interferes with previous knowledge: not only the exact features of a perceived object are linked together via inter-area synapses, but also similar features which lie inside the activation bubble, and so are simultaneously co-active.

Simulation results, obtained by using a simple network with a minimum of internal complexity, and using an abstract representation of objects (as a collection of 4 features) demonstrate that the proposed mechanism may actually work, producing a high percentage of success (more than 90%). Moreover, simulation results provide some interesting indications on the ease or difficulty to recognize multiple objects, which, if confirmed on subsequent more physiological models, may represent the subject for future validation studies via psychophysical tests.

A first important aspect of our model is the possibility to recognize objects starting from an initial incomplete representation. This mechanism in part resembles that exploited in auto-associative memories [19]. Another important point in our model is the trade-off between sensitivity and specificity. In the model, this trade-off can be managed acting on the extension of lateral synapses, that is on the dimension of the activation bubble. Sensitivity means the capacity to detect an object even in difficult conditions, i.e., if some features are absent or corrupted from the original ones (that is the capacity to detect true positives by reducing false negatives). Specificity means the capacity to discriminate one object from a similar one, by maintaining the two representations separate (i.e., the capacity to avoid false positives still rejecting true negatives). Of course, good sensitivity may be associated with poor specificity, and vice versa.

In our model, if the activation bubble is small (as in the simulations summarized in Table 1) the network

ISSN: 1727-1983 © 2005 IFMBE

exhibits a good compromise between sensitivity and specificity. It is able to reconstruct objects if one feature is lacking and another feature is corrupted by one position, while an object is not reconstructed if a property is corrupted by two positions. A greater sensitivity may be achieved by extending the lateral excitatory synapses (unpublished simulations) but worsening specificity.

Conclusions

The present model represents a first attempt to achieve object reconstruction and segmentation, by implementing high-level Gestalt rules within the framework of neural synchronization in the γ-band. Results show that the network is able to reconstruct partially corrupted objects, and to reject objects with a higher level of corruption or incompleteness, with a high percentage of success (greater than 90%) and with acceptable settling times (30-40 ms). At present, the model does not aspire to reflect neurophysiological or neuroanatomical knowledge in detail, but rather to propose a computational mechanism, which exploits and extends some current ideas on memory and learning [4,5,18,22]. Future lines may be directed both toward an improvement of the computational aspects (i.e., the capacity to recognize objects in different conditions with a flexible and reliable performance) or toward a more precise connection with neurophysiology and neuroanatomy.

References

[1] SINGER W. (2002): 'Synchronization, binding and expectancy', in ARBIB, M. A. (Ed): 'The handbook of brain theory and neural networks', (The MIT Press, Cambridge, MA), pp. 1136-43

[2] BARLOW H. B. (1972): 'A neurone doctrine for perceptual psychology', *Perception*, **1**, pp. 371-394

[3] SINGER W. (1999): 'Neuronal synchrony: a versatile code for the definition of relations?', *Neuron*, **24**, pp. 49-65

[4] VON DER MALSBURG C. and SCHNEIDER W. (1986): 'A neural cocktail-party processor', *Biol. Cybern.*, **54**, pp. 29-40

[5] VON DER MALSBURG C. and BUHMANN J. (1992): 'Sensory segmentation with coupled neural oscillators', *Biol. Cybern.*, **67**, pp. 233-246

[6] SINGER W. (1993): 'Synchronization of cortical activity and its putative role in information processing and learning', *Annu. Rev. Physiol.*, **55**, pp. 349-374

[7] SINGER W. and GRAY C. M. (1995): 'Visual feature integration and the temporal correlation hypothesis', *Ann. Rev. Neurosci.*, **18**, pp. 555-586

[8] ECKHORN R. (1999): 'Neural mechanisms of scene segmentation: recordings from the visual cortex suggest basic circuits for linking field models', *IEEE Trans. Neural Net.*, **10**, pp. 464-479

[9] GRAY C. M. and SINGER W. (1989): 'Stimulus-specific neuronal oscillations in orientations columns of cat visual cortex', *Proc. Natl. Acad. Sci. USA*, **86**, pp. 1698-1702

[10] TOVEE M. J. and ROLLS E. T. (1992): 'Oscillatory activity is not evident in the primate temporal visual cortex with static stimuli', *NeuroReport,* **3**, pp. 369-372

[11] TALLON-BAUDRY C., BERTRAND O., PERONNET F., and PERNIER J. (1998): 'Induced gamma-band activity during the dealy of a visual short-term memory task in humans', *J. Neurosci.*, **18**, pp. 4244-4254

[12] ENGEL A. K., FRIES P., and SINGER W. (2001): 'Dynamic predictions: oscillations and synchrony in top-down processing', *Nat. Rev. Neurosci.,* **2**, pp. 704-716

[13] KAISER J., LUTZENBERGER W., ACKERMANN H., and BIRBAUMER N. (2002): 'Dynamics of gamma-band activity induced by auditory pattern changes in humans', *Cereb. Cortex*, **12**, pp. 212-221

[14] KOFFKA K. (1935): 'Principles of Gestalt Psychology', New York: Harcourt

[15] WANG D. and TERMAN D. (1997): 'Image Segmentation Based on Oscillatory Correlation', *Neural Computation*, **9**, pp. 805-836.

[16] LI Z. (1998): 'A neural model of contour integration in the primary visual cortex', *Neural Computation*, **10**, pp. 903-940

[17] URSINO M., LA CARA G. E., and SARTI A. (2003): 'Binding and segmentation of multiple objects through neural oscillators inhibited by contour information', *Biol. Cybern.*, **89**, pp. 56-70

[18] WANG D., BUHMANN J., and VON DER MALSBURG C. (1990): 'Pattern segmentation in associative memory', *Neural Computation*, **2**, pp. 94-106

[19] HERTZ J., KROGH A., and PALMER R.G. (1991): 'Introduction to the theory of neural computation', Redwood, CA: Addison-Wesley

[20] RODRIGUEZ E., GEORGE N., LACHAUX J. P., MARTINERIE J., RENAULT B., and VARELA F. J. (1999): 'Perception's shadow: long-distance synchronization of human brain activity', *Nature,* **397**, pp. 430-433

[21] GRUBER T., MÜLLER M. M., and KEIL A. (2002): 'Modulation of induced gamma band responses in a perceptual learning task in the human EEG', *J. Cogn. Neurosci.,* **14**, pp. 732-744

[22] KAISER J. and LUTZENBERGER W. (2003): 'Induced gamma-band activity and human brain function', *The Neuroscientist*, **9**, pp. 475-484

ISSN: 1727-1983 © 2005 IFMBE