

# Towards Word Semantics from Multi-modal Acoustico-Motor Integration: Application of the Bijama Model to the Setting of Action-Dependant Phonetic Representations

Olivier Ménard<sup>1,2</sup>, Frédéric Alexandre<sup>2</sup>, and Hervé Frezza-Buet<sup>1</sup>

<sup>1</sup> Supelec, 2 rue Edouard Belin, 57070 Metz, France

<sup>2</sup> Loria, Campus Scientifique, BP 239, 54506 Vandoeuvre-ls-Nancy, France

**Abstract.** This paper presents a computational self-organizing model of multi-modal information, inspired from cortical maps. It shows how the organization in a map can be influenced by the same process occurring in other maps. We illustrate this approach on a phonetic - motor association, that shows that the organization of words can integrate motor constraints, as observed in humans.

## 1 Introduction

In the evolutionary process, the appearance of the cerebral cortex has had dramatic consequences on the abilities of mammals, which reach their maximum in humans. Whereas it can be said that the limbic system has added an emotional dimension on purely reactive schemes [1], the cerebral cortex has offered a new substratum devoted to multimodal information representation [2]. When one considers the associated cost of this neuronal structure in terms of energy needs and size in the limited skull, it can be thought that the corresponding functions might be complex but highly interesting from an adaptive point of view.

Basically, the cerebral cortex is often described as a set of topological maps representing sensory or motor information, but also merging various representations in so-called associative maps. Concerning afferent connections toward cortical maps, the topological principle explains that information coming from sensors is represented along important dimensions, like retinotopy for the visual case. Moreover, at a lower level of description, some kind of filtering process allows to extract and represent onto the mapping other functional information [3], like orientation selectivity or color contrast in the visual case.

Concerning cortico-cortical connections, the important role of these internal links must be underlined. For example, the cerebral cortex and the cerebellum are reported as having approximately the same number of synapses ( $10^{12}$ ) [4] and the big difference of volume between these structures can be explained by the fact that internal connections are much more numerous in the cerebral cortex (more than 75%). These internal connections inside the cerebral cortex are observed as belonging to a map, to achieve the topological representation, but also as connecting maps, which is fundamental to create associative maps [5, 2].

From a functional point of view, the role of the cerebral cortex has often been described as unsupervised learning [6]. In the statistical domain, the goal of unsupervised models like the K-means, hierarchical classification, Principal Component Analysis (PCA), Independent Component Analysis (ICA) is to categorize information from the regularities observed in its distribution (as opposed to an external signal, seen as a teacher or a supervisor) or to select in a high dimensional space the most significant axes on which to project information. It must be underlined that such information processing is very consistent with the cortical organizational principles of topological representation and filtering.

From a modeling point of view, neuronal models are among the most well-known unsupervised techniques. The central one is certainly Kohonen's Self-Organizing Map [7], which has been proposed from its origin as a model of a cortical map (see also [8]) and has been applied in various sensory domains (see for example [9] for the visual case, [10] for the auditory case, etc.). Later, from this simple but powerful scheme, other more complicated models have been elaborated to fit more closely to the biological reality (cf. for example [11] for the visual case), but they all rely on the same fundamental principle of competitive learning, as observed in the cerebral cortex.

Interestingly, it must be noticed that most of these neuronal models lay emphasis on the representation of one sensory or motor information and not on the joint organization of several interacting flows of information (a notable exception being [12]). Nevertheless, evidence from neurosciences indicates that this function is also present in cortical processing. To tell it differently, the cortex is not only several self-organizing maps, each one representing its own modality (or set of modalities in the associative case) and communicating one with the other, but rather a set of maps acting all together to represent information of the external world from different but cooperating points of view in a global way.

Of course, such a holistic view cannot be obtained if, as it is often the case, one unique map is considered in the modeling process. The fact is that several biological data indicates that the cortical processing cannot be only summarized by independent self-organizations.

From a connectivity point of view, we have indicated above the important role which is given to recurrent cortico-cortical connections. This might be consistent with asking for a global consistency in representations on top of simple local competitions. Several electrophysiological studies have shown that a cortical region can change the kind of information it represents in case of a lesion (e.g. changing representation of the finger in a lesioned monkey [13]) or in case of sensory substitution (e.g. tactile stimulation for blind people [14]).

From a representational point of view, several brain imaging studies [15] have shown that word encoding within the brain is not only organized around phonetic codes but is also organized around action.

How this is done within the brain has not yet been fully explained but we would like to present how these action based representations naturally emerge in our model by virtue of solving constraints coming from motor maps. This model, called Bijama, is a general-purpose cortically-inspired computational framework

that has also been used for rewarded arm control [16]. It is described in general terms in section 2, with particular attention to the effect of learning rules in section 3. Then, the actual use of the model for acoustico-motor integration is presented in section 4.

## 2 Bijama Model Features

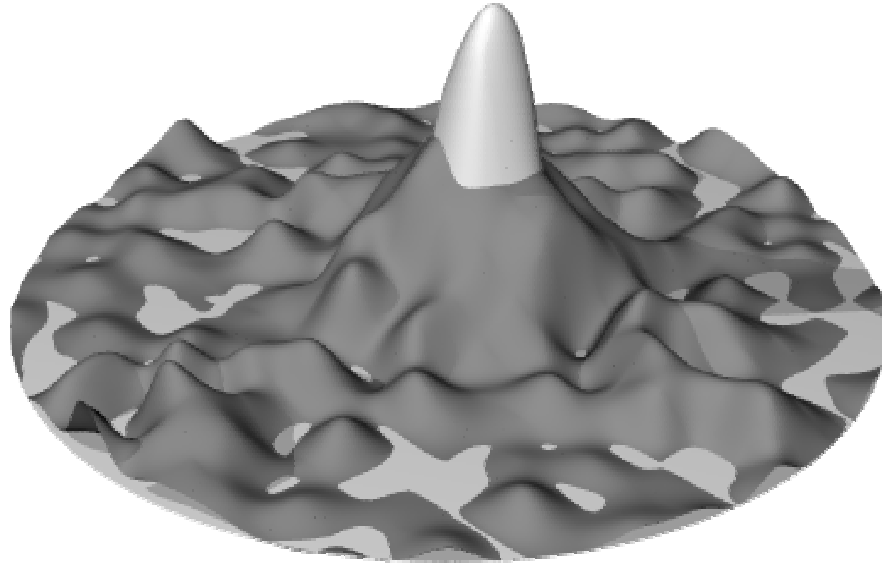
The features of the model are presented briefly in the following sections. A more detailed presentation can be found in [17, 16], where the current model is applied to a simplified version of a target reaching problem with an artificial arm. This model is referred as the Bijama model in related papers, which stands for Biologically-Inspired Joint Associative MAs.

### 2.1 Maps, Units and Competition

The main computational block of the model is a set of computational units called a map. A map is a sheet made of a tiling of identical units. This sheet has been implemented as a disk, for architectural reasons described further. When input information is given to the map, each unit shows a level of activity, depending on the similarity of the information it receives with the information it specifically detects, as will be detailed in section 2.2. That activity, noted  $A^t$ , follows a Gaussian *tuning curve* in the model:  $A^t$  is a matching activity, that is maximal if input information exactly corresponds to the *prototype* of the unit, and gets weaker as input gets different from this prototype.

When an input is given to the map, the distribution of matching activities among units is a scattered pattern, because tuning curves are not sharp, which allows many units to have non null activities, even if prototypes don't perfectly match the input. From this activity distribution over the map, a small compact set of units that contains the most active units has to be selected. Unlike in SOMs where this decision is made by a centralized "winner-take-all" process, decision is made here by a numerical distributed process, emerging from a local competitive mechanism, as in [8].

In order to decide which units are locally the best matching ones inside a map, a local competition mechanism is implemented. It is inspired from theoretical results of the continuum neural field theory (CNFT) [18, 19], but it is adapted to become independent of the number of connections, thus avoiding disastrous border effects: The CNFT algorithm tends to choose more often units that have more connections. Thus, the local connection pattern within the maps must be the same for all units, which is the case for torus-like lateral connection pattern, with units in one border of the map connected to the opposite border, for example. Here, the field of units in the map computes a distribution of global activities  $A^*$ , resulting from the competition among current matching activity  $A^t$ . This competition process has been made insensitive to the actual position of units within the map, in spite of heterogeneous local connection patterns at the level of border units in the Bijama model, as detailed further.



**Fig. 1.** Result of competition in a map among the  $A^t$  (dark gray) and resulting  $A^*$  (light gray). A bubble of  $A^*$  appears where  $A^t$  is the strongest in its neighborhood.

The result of this competition is the rising of a bubble of  $A^*$  activity in the map at places where  $A^t$  activities are the most significant (cf. figure 1). The purpose of the resulting  $A^*$  activity is twofold. First, this activity defines the main activity of the unit: This activity is the one that is viewed by other connected units in all activation rules detailed further. Second, all learning processes are modulated by this activity. That means that only units in  $A^*$  activity bubbles learn in the map.

The global behavior of the map, involving adaptive matching processes, and learning rules dependent on a competition, reminds the Kohonen SOM. However, the local competition algorithm used here allows the units to be feed with different inputs. The source of information received by a unit differs from one unit to its neighbors, because of the stripe connectivity described below in section 2.3. Another difference with SOM not previously detailed is that, in our model, competition and learning are not separated stages. Learning is dependent on  $A^*$ , and also occurs during the  $A^*$  bubble setting.

## 2.2 Matching Activity Computation

It has been mentioned previously that competition is computed on the basis of a matching activity  $A^t$ . As detailed below, this activity is actually the merging of several matching results, and it may be considered as a global matching activity. Inside the units in the model, each matching result is performed by a

computational module called a layer. Therefore, a layer in our model is a sub-part of a unit, computing a specific matching, and not a layer of neurons as classically reported in various models. It is inspired from the biological model of the cortical column by [20]. A layer gathers inputs from the same origin (a map), and computes a matching value from the configuration of these inputs. As a consequence, the behavior of a unit can be described as the gathering of several layers. These are detailed in the following.

First of all, some maps receive input from the external world. Each unit in the map reacts according to the fitting of this input to a preferred input. In the cortex, the thalamus plays a role in sending inputs to the cortex. In our model, the layer which tunes a preferred perception is called a thalamic layer. This layer provides a thalamic matching activity.

One other kind of layer is the cortical layer. It receives information from another map. The connectivity of this layer will be further discussed in section 2.3. Let us just say for now that its purpose is to compute a cortical matching activity that corresponds to the detection of some  $A^*$  activity distribution in the remote units it is connected to.

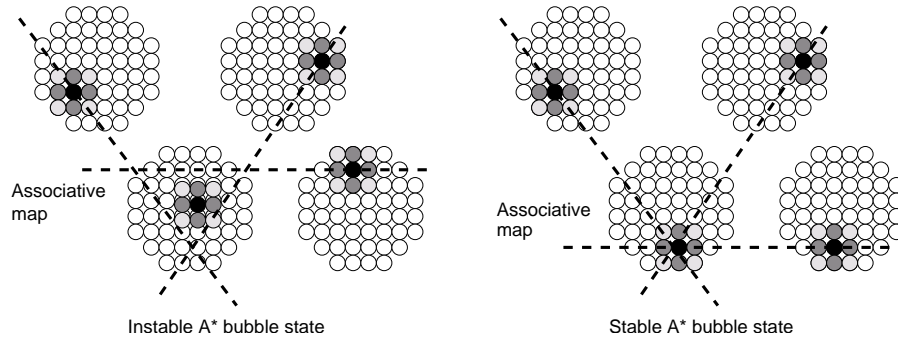
If the map is connected to a number  $n$  of other maps, its units have  $n$  cortical layers, thus computing  $n$  cortical matching results (one per cortical layer). These matchings are merged to form a global cortical matching. If the map has a thalamic layer, the thalamic matching result is then merged to the global cortical matching, to form the global matching  $A^t$  the competition is performed from.

To sum up, our model stresses the two kinds of cortico-cortical connections mentioned in section 1: Local connections and inter-map connections. The maps compute activity bubbles, that are a decision enhancing the most relevant units from local connections belonging to the map. This decision depends on external input, computed by the thalamic layer, but also on the state of other maps through cortical layers, that implement long range cortico-cortical connections. This computation is a multi-criteria decision, that has complex dynamics, since it performs a competition from input, but also from the competition that is performed in the same way in other maps. One consequence of this dynamics, central to the model, is that the self-organization in a map is modulated by the organization in the other maps, as illustrated in section 4.

### 2.3 Inter-map Stripe Connectivity and Disk-Shaped Maps

A cortical layer, that receives information from another map, doesn't receive inputs from all the units of the remote map, but only from one stripe of units (cf. fig. 3). For instance, a map may be connected row-to-row to another map: Each unit in any row of the first map is connected to every remote units in the corresponding row of the other map. These connections are always reciprocal in the model.

This limited connectivity is biologically grounded, as cortical zones are connected to other zones by stripes [2, 3]. Moreover, it has a computational purpose: If inter-map connectivity were total (if each unit in a map were connected to every unit in a connected remote map), the number of connections would rise



**Fig. 2.** Conditions for stable  $A^*$  activity states. On the left, instable state where bubbles of  $A^*$  don't stand at intersecting stripes. On the right, stable state: Any other bubble position in the same stripes on the three non-associative maps would also be stable.

too quickly as the size and the number of the maps increase and would lead to a combinatorial explosion. Since this model has been designed to handle multiple sensori-motor connections, the risk is real and map-to-map connectivity has to be limited.

A stripe has an orientation that is specific to a map-to-map connection: A map that is connected to many other ones has a different direction of connection for each kind of connection. In order to keep the model homogeneous, the shape of the map must not favor any direction. This is the reason why the maps are disk-shaped in our model.

Two analogous cortical layers of two neighboring units are connected to parallel, adjacent and overlapping stripes in the remote map. Neighboring units receive close but not identical inputs. That is why a winner-takes-all algorithm over the whole map isn't suitable, as already explained.

Through the inter-map connectivity, our model produces *resonance* between connected maps: Activity patches in connected maps can only stabilize within connected modular stripes. The role of reciprocally connected stripes is crucial for this resonance. Activity  $A^t$  is the basis for inner-map lateral competition (computation of  $A^*$ ). As this  $A^t$  depends on some cortical activity(ies), computed from other cortical inputs that are fed with remote  $A^*$ , bubbles of  $A^*$  activities raise in the maps at some particular places: The bubble of activity that appears in an associative map is at the intersection of the stripes where activity bubbles coming from the connected maps stand (see figure 2).

In our model, this matching of activity can be compared with a phenomenon of resonance, as described in the ART paradigm by Grossberg [21], that produces stable and coherent states across the different maps. It ensures consistency of the activity bubbles across two connected cortical maps. Since units learning rate is modulated by their  $A^*$ , units whose  $A^*$  are activated simultaneously in the different maps learn together. We call this *coherent learning*. Learning

strengthens the connection between these coherent units, so that they will tend to activate together again in the future.

## 2.4 Activation and Learning Rules

As mentioned before, cortical and thalamic layers of the units in the model have to perform a tuning from the input they receive, so that all matchings are merged to constitute the global matching activity  $A^t$ . This merging concerns all cortical and thalamic layers, and is computed from a geometric mean. This must be seen as a tricky way to compute some kind of numerical AND operator. Knowing these merging principles, let the computation of each elementary matching, and their associated learning rule, be detailed for both thalamic and cortical layers.

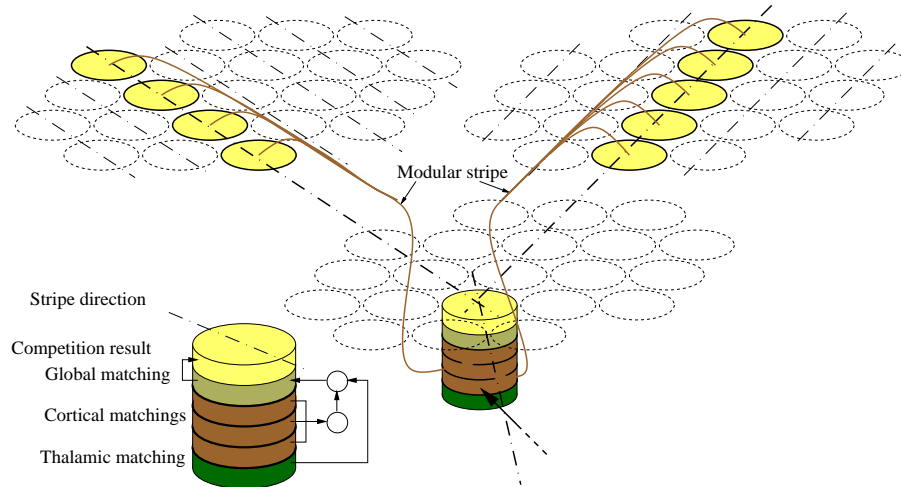
The thalamic layer in the model behaves similarly to neurons in Kohonen maps. This is a custom defined point in the model, depending on the actual entry format received by the map. For example, thalamic tuned activation can be a decreasing function of a well suited distance between the input and a prototype. Then learning consists of making the thalamic prototype be closer to the current input. This learning process has to be modulated by  $A^*$  activity for thalamic layer to be coherent with the remaining of the model. This is also what is done in Kohonen maps, where learning rate depends on a decreasing function of the proximity of a neuron with the winning one. This decreasing function in Kohonen algorithm is analogous to the  $A^*$  bubble of activity in the model.

The cortical layers all use the same matching and learning rules. Each cortical activity is computed from a cortical prototype pattern and the cortical input pattern, which is actually the  $A^*$  activity distribution in the connected stripe of remote units. The layer matching activity has to be high only when both the  $A^*$  of a remote unit and the corresponding value in the prototype are high: The cortical layer detects that a remote unit to which it is highly connected is active, and thus performs a computational “AND”. The learning is, as for the thalamic layer, modulated by  $A^*$ .

A unit learns only when it actively participates in the recognition process, i.e. when it is at a place where a  $A^*$  bubble stands. It learns both its thalamic and cortical prototypes, which creates and then maintains coherence between the different layers. The full unit model is summarized in figure 3.

## 2.5 Joint Organization

To conclude on the model behavior, the combination of self-organization and coherent learning produces what we call *joint organization*: Competition, although locally computed, occurs not only inside any given map, but across all maps. Moreover, the use of connection stripes limits the connectivity, which avoids the combinatorial explosion that would occur if the model were to employ full connectivity between the maps. Thus, coherent learning leads to both efficient data representation in each map and coordination between all connected maps.



**Fig. 3.** Full functional scheme: The cortical matching activities, obtained from the modular stripe inter-map connections, are merged together. The thalamic matching is merged with the result to form a global matching activity. This activity is then used in the competition process described in section 2.1

### 3 Multi-association and Cortical Learning Rules

A multi-associative model is intended to associate multiple modalities, regardless of however they are related. It must then handle the case where the associations between two modalities are not one-to-one, but rather one-to-many, or even many-to-many. This *multi-association* problem will now be presented on a simple example.

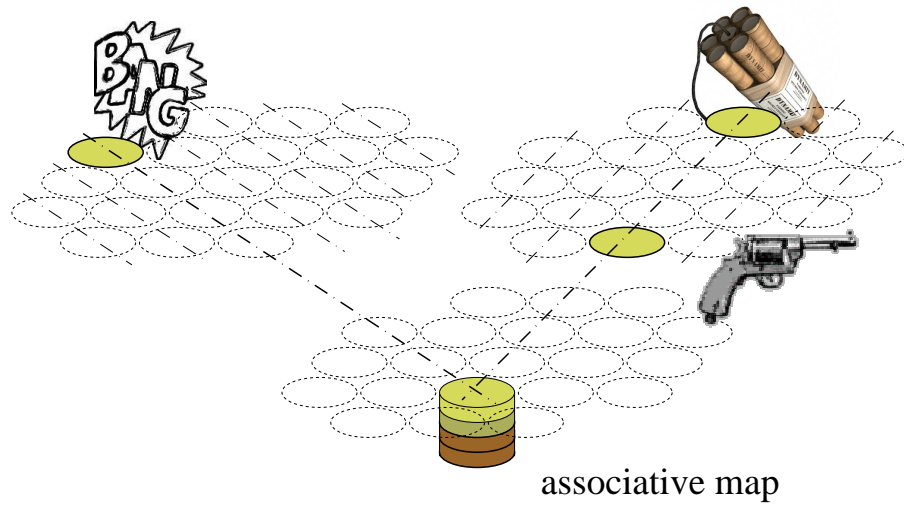
#### 3.1 Associative units and Multi-association

Let us consider an association between certain objects and the sounds they produce. A car, for example, could be associated with a motor noise. Certain objects produce the same noise. As a result, a single noise will be associated with multiple objects. For instance, a firing gun and some exploding dynamite produce basically both an explosion sound.

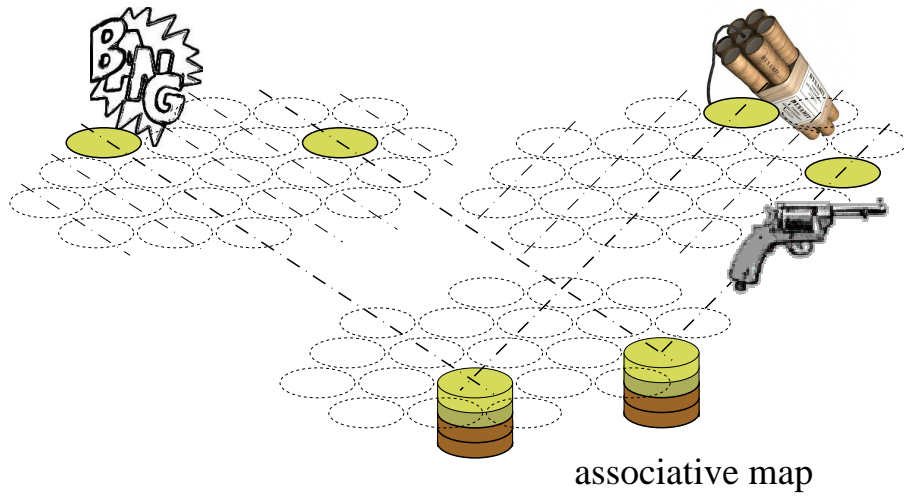
In our model, let us represent the sounds and the objects as two thalamic modalities on two different cortical maps. Let us now link both of these maps to another one, that we call an *associative* map. The sound representations and the object representations are now be *bound* together through the associative map (see fig. 4).

If we want a single unit to represent the “BANG” sound in the sound map, a single unit in the associative map has to bind together the “BANG” unit with both the gun and the dynamite units. This associative unit must then have the





(a)



(b)

**Fig. 4.** A multi-association: A sound is associated with two different object, both of which may produce this actual sound. (a) A single sound unit is connected with the two object units by a unit in the associative map, that stands at the intersection of two sound and object stripes: This is possible because the unit's global connection strength is not distributed among all connections, but only among *active* connections (Widrow-Hoff rule). (b) When a unit's global connection strength is distributed among all connections, the sound unit must be duplicated (Hebb rule). See text for detail.

ability to perform *multi-associations*: It must have a strong cortical connection to two different units (the gun and the dynamite units) in the same cortical stripe (see fig. 4a).

If associative units cannot perform multi-associations, the resulting self-organization process among all maps will duplicate the “BANG” representative unit. The reason is that, in this very case, an associative unit is able to listen to only one unit in each connected module. Each instance of that unit will then be bound, through its own associative unit, either to the dynamite or the gun unit (see fig. 4b). Moreover, the two object units cannot be in the same connectivity stripe, or else the model would try to perform multi-association and fail.

In fact, in the associative map, a unit is able to bind all possible couples of sound and object units that are in the two cortical stripes. Reciprocally, two given stripes of sound and object units can only be bound through a single associative unit in the associative map, the one that stands at the intersection of these stripes. Therefore, since our model uses a logical “AND” between the different inputs of a unit, that single associative unit must be active each time one of these sound and one of these objects are active together.

Thus, if a unit can only handle one association, each couple of sound and object stripes must contain, at most, one associated object and sound couple of units. The units that represent different object making the same sound, for instance the dynamite and the gun, must then be on different cortical stripes (see fig. 4a). The same is true for the two “BANG” units, which represent the same sound. Actually, the ability of units to handle multi-associations depends solely on the cortical learning rule used in the model, as we will now explain. This ability is crucial in the model, as explained further.

### 3.2 Cortical Learning Rules Requirements for Multi-association

It is the cortical learning rule that enables, or not, a unit to be strongly connected to many remote units in the same cortical stripe. Therefore, the cortical learning rule is the key to solving the multi-association problem.

Let us consider first the Hebb/anti-Hebb learning rule. Using this rule, if unit  $i$  is connected to unit  $j$ , the weight  $w_{ij}$  of the cortical connection from  $i$  to  $j$  is updated through :

$$\delta w_{ij} = \delta A_i^* \times (A_j^* - w_{ij})$$

where  $\delta$  is the update rate of the cortical connection weights. Thus, the connection strength  $w_{ij}$  between the local unit  $i$  and a remote unit  $j$  grows when both  $i$  and  $j$  are active together (i.e. they are both in an  $A^*$  activity bubble).  $w_{ij}$  decreases each time the remote unit  $j$  is inactive while the local unit  $i$  is active.

Therefore, if unit  $i$  and  $j$  are *always* active together,  $w_{ij}$  will grow to reach the maximum weight. Consider now the case where unit  $j$  is active, for instance, half of the time when unit  $i$  is active, and remote unit  $j'$  is active the other half of that time. Both  $w_{ij}$  and  $w_{ij'}$  will only reach half the maximum weight. Since

cortical activation is computed on the base of the  $w_{ik} \times A_k^*$ , the cortical input from  $j$  (or  $j'$ ) can be quite weak when  $A_j^*$  (or  $A_{j'}^*$ ) is strong, just because many  $A_{j'}^*$  were correlated to the  $A_i^*$  activity.

Since only the most active units are inside the  $A^*$  activity bubble, a local unit  $i$  can be active only if its cortical activities are high enough. Therefore, because of competition, local unit  $i$  can only be active if its connection weight to some remote unit  $j$  is high. This is the reason why, using the Hebb/anti-Hebb learning rule, a unit  $i$  can be activated by a unit  $j$  if  $A_j^*$  is the only one that is correlated to  $A_i^*$ .

This result is actually due to the following: The global connection strength of the local unit  $i$  for a given cortical stripe is at every time distributed among all connected remote units  $j$ . Since that connection strength must be high for some connection, it is concentrated on a single remote unit, which means that all other remote units are very weakly connected to the local unit. The end result is that the local unit binds together a single remote unit per stripe connection.

As a consequence, the model cannot represent a situation where a unit in a map should be bound with multiple units in the other remote maps: It cannot handle *multi-associations*. The only way to effectively associate a given thalamic input in a map to two different thalamic inputs in another map is to have it represented by two different patches units (see fig. 4b).

This can be avoided if the cortical learning rule allows a unit to be strongly associated with multiple units for each cortical stripe. A cortical learning rule that allows this is the Widrow-Hoff learning rule.

### 3.3 Widrow-Hoff Learning Rule and Consequences for Multi-association

Using a learning rule adapted from the Widrow-Hoff learning rule, if unit  $i$  is connected to unit  $j$ , the weight  $w_{ij}$  of the cortical connection from  $i$  to  $j$  is updated through :

$$\delta w_{ij} = \delta(A_i^* - \omega) \times (A_i^* - A_i^c) \times A_j^*$$

where  $\delta$  is the update rate of the cortical connection weights,  $A_i^c$  is the cortical activity of a unit  $i$ , and  $\omega$  is the decay rate of cortical connections. Here, cortical activity  $A_i^c = \sum_j w_{ij} A_j^*$  is seen as a predictor of  $A_i^*$ . When both the local unit  $i$  and the remote unit  $j$  are active together, if  $A_i^c$  is lower than  $A_i^*$ ,  $w_{ij}$  grows, and if  $A_i^c$  is higher than  $A_i^*$ ,  $w_{ij}$  decreases.  $w_{ij}$  also decreases slowly over time.

Here, the global connection strength of the local unit  $i$  for a given cortical stripe is distributed among all *active* remote units  $j$ , and *not* among *all* remote units. As with the Hebb/anti-Hebb rule, because of the local competition, the connection strength  $w_{ij}$  between  $i$  and a unit  $j$  must be high. However, here, raising  $w_{ij}$  doesn't imply lowering all  $w_{ik}$  for all  $k$  in the remote connection stripe. Raising  $w_{ij}$  will only lower  $w_{ik}$  if  $j$  and  $k$  are active at the same time.

However, since only a small  $A^*$  activity bubble is present on each map, most remote units in the connection stripe cannot be active at the same time. Thus,

the local unit  $i$  can bind together multiple units in a given stripe connection to a unit in another stripe connection. This is the reason why the use of the Widrow-Hoff learning rule in our model leads to multi-map organization as in fig. 4a.

Solving the multi-association problem has one main benefit: The maps need fewer units to represent a certain situation than when multi-association between unit is impossible. Moreover, since instances of a given thalamic input are not duplicated in different parts of a cortical map, it is easier for the model to perform a compromise between the local organization and the cortical connectivity requirements, i.e. *joint organization* is less constrained.

## 4 Model behavior on a simplified example

### 4.1 The Phonetic-Action Association Problem

Several brain imaging studies [15] have shown that word encoding within the brain is not only organized around purely phonetic codes but is also organized around action. How this is done within the brain has not yet been fully explained but we would like to present how these action based representations naturally emerge in our model by virtue of solving constraints coming from motor maps.

We therefore applied our model to a simple word-action association. A part of the word set from the European MirrorBot project, which is a 3 year EU-IST-FET project, was used in a “phonetic” map, and we tried to associate these words to the body part that performs the corresponding action. One goal of this project is to define multimodal robotic experiments and the corresponding protocols are consequently well suited for this task.

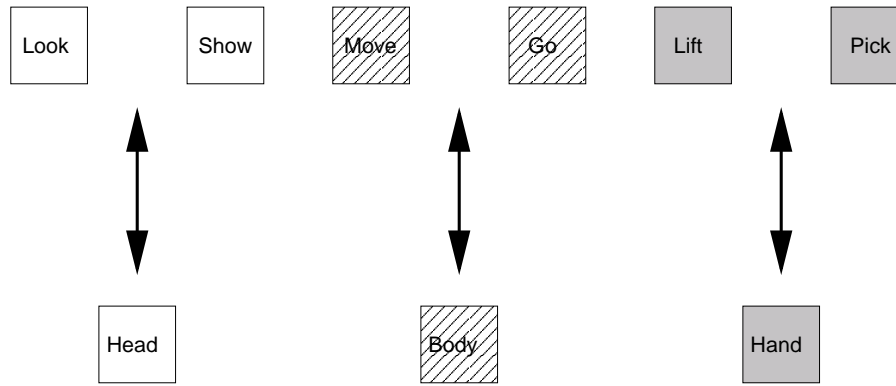
### 4.2 Phonetic and Motor Coding

The phonetic coding used in our model is taken from the MirrorBot project. A word is separated into its constituting phonemes. Each phoneme is then coded by a binary vector of length 20. Since the longest word that is used has 4 phonemes, each word is coded by 4 phonemes, and if they have less, they are completed by empty phonemes.

The distance between two different phonemes is the Cartesian distance between the coding vectors. The distance between two words is the sum of the distances between their constituting phonemes. While we are well aware that this is a very functional way to represent the phonetic distance between two words, it is sufficient in order to exhibit the joint organization properties discussed in this paper.

The actions are coded in the same way as the words: There are 3 different input actions (head action, body action and hand action), and each action is coded as a binary vector of length 3. The distance between two actions is, once again, the Cartesian distance between their representing vectors.

Each word is semantically associated to a specific action. The word-action relationship is shown on figure 5.



**Fig. 5.** Word-Action Relationship

The thalamic prototypes (i.e. external inputs) of the motor and the phonetic units are, respectively, coded actions and coded words. However, these do not necessarily correspond to real input words or actions: These prototypes are vector of float values, not binary ones. The prototype of a unit, in the figures of this section, is represented as the nearest “real” input, in term of the distance previously discussed.

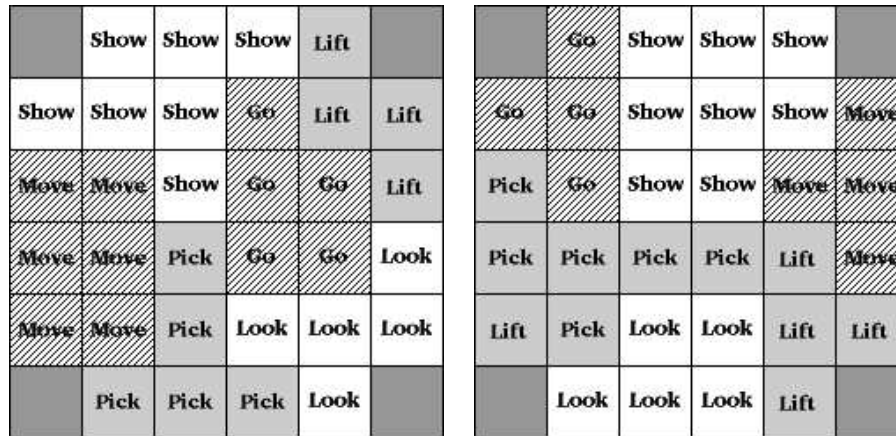
### 4.3 Interest of Associative Learning

Our model fundamentally differs from a classical Kohonen map since this latter one is somehow topologically organizing information against the sole notion of distance between inputs and prototypes. Thus if we were to use a Kohonen map to represent words from the MirrorBot grammar (encoded as a phonetic sequence), a consequence of the Kohonen algorithm and existing lateral interaction between units would be an organization toward similarity relation of word codes only (i.e. two words having similar code would be represented by the same prototype or neighbor prototypes) as illustrated in figure 6. This kind of representation is not satisfactory in the sense that it is totally disconnected from other maps and does not take any semantics of words into account.

### 4.4 Emergence of Action Oriented Representation

Let us consider three maps, one for word representation, one for action representation and finally an associative one that links word to action (cf. figure 7).

The central point of our model is that coherent learning within a map depends on some other maps, so that the inter-map connectivity biases the convergence to a particular self-organized state, when self-organization alone would have allowed for many more possible ones. The final state of organization in each map must allow the bubbles to be set up at intersecting cortical connection



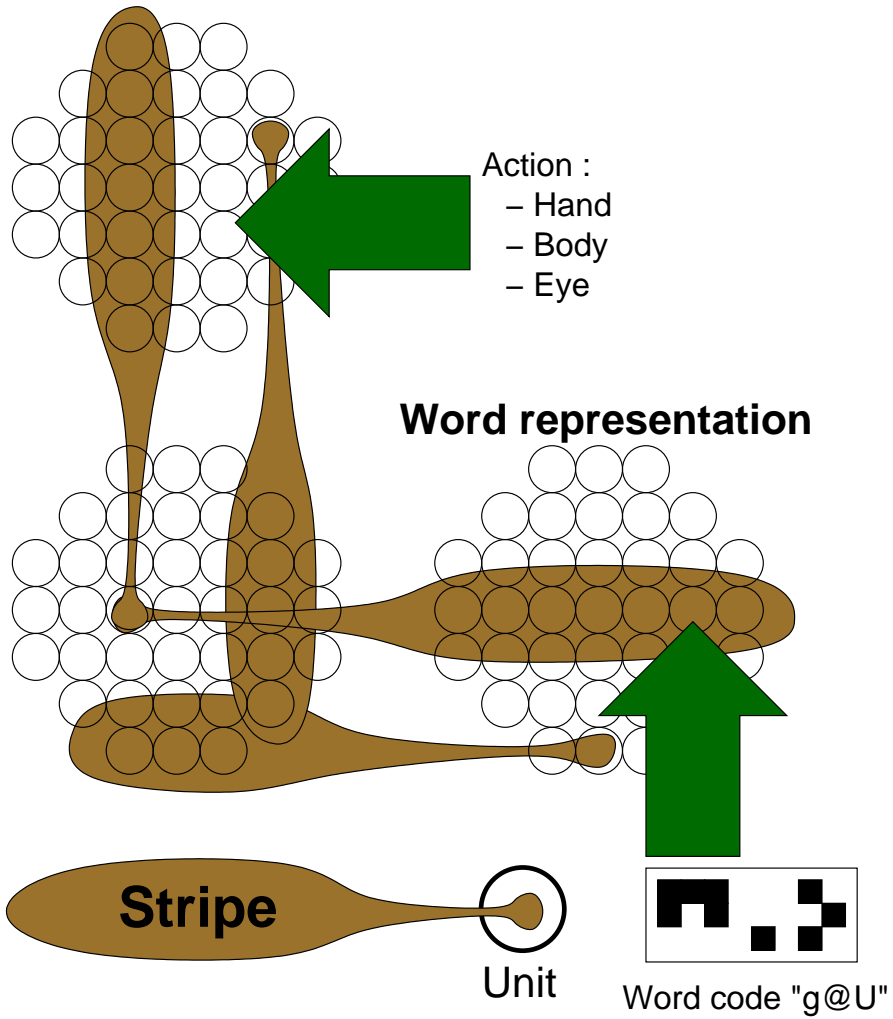
**Fig. 6.** Two different results of word classification by a Kohonen map based on purely phonetic representations. Words representing eye action (white), hand action (gray) or body action (stripes) are spread all over the map without paying any attention to the underlying semantic of words.

stripes, solving inter-map constraints as the one illustrated on fig 2. The cortical maps perform an effective compromise between the local and remote constraints. Remote constraints, coming from the architecture, makes activity bubbles have strong cortical connections to each other. Local constraints, coming from the thalamic layers, requires bubbles of activity to raise where the phonetic or action prototypes best match the phonetic or action input. This compromise is poor at the beginning, but it gets better as learning proceeds.

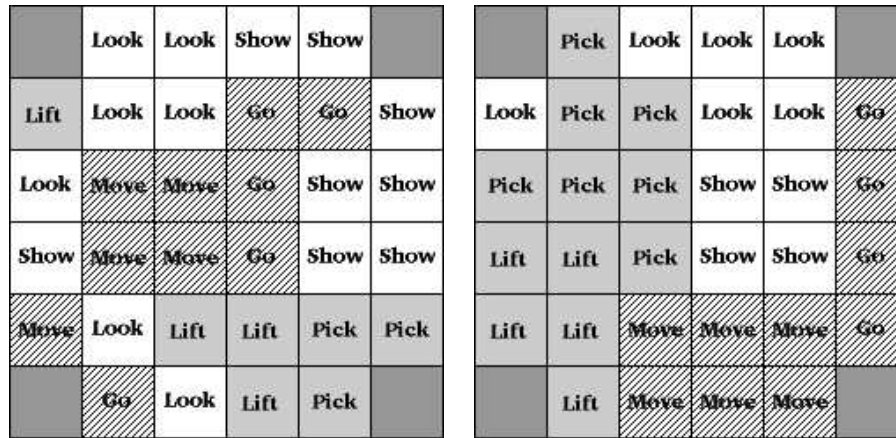
In the current word-action association, we have chosen to impose a frozen organization to the action map, in order to illustrate how the phonetic map self-organizes when keeping coherence with the action map. As an illustration, let us consider the words “look” and “show”. The phonetic representations of these words are completely different, so that a Kohonen map classifies them in different parts of the map (cf. fig. 6). In our model, however, the higher level associative map linking auditory representation with motor action will use close units to represent these words, since they both relate to the same action (head action), see “look” and “show” positions on fig. 8. As our model deals with an implicit global coherence, it is able to reflect this higher level of association and to overcome the simpler phonetic organization.

The interesting point to consider here is that word representations (e.g. phonetic map) are constrained by some topology that mimics to some extent physical properties of effectors, i.e. a word unit is dedicated to one action (e.g. hand) and cannot trigger another one (e.g. head). In order to solve this constraint and to ensure a global coherence, the model must then organize word representation

## Action representation



**Fig. 7.** Schematic view of the model architecture: The word representations and the action representations are presented in separate maps, that are both connected to an associative map by reciprocal cortical connection stripes.



**Fig. 8.** Two simulation results of word representation map after coherent learning has occurred with our model. Word representations are now constrained by the motor map via the associative map and, as a result, words that correspond to the same action are grouped together. Nevertheless, phonetic proximity is still kept.

in such a way that, for example, any “body” word should be linked to a body action.

As illustrated in figure 8, we can clearly see that the topological organization found by the model meets these criteria. Within the word map, words are grouped relatively to the body part they represent: Body action words are grouped together (stripes) as well as hand action words (gray) and head action words (white).

However, the phonetic distribution of words remains the most important factor in the phonetic map organization. Each word is represented by a “cluster” of close units, and the words whose phonetic representation is close tend to be represented in close clusters of units. For instance, while “Go” and “Show” correspond to different motor actions, their phonetic representations are close, so that their representing clusters are adjacent (cf. fig. 8). This illustrates the fact that the model is actually doing a successful compromise between the local demands, which tend to organize the words phonetically, and the motor demands, which tend to put together the words that correspond to the same action. The joint organization does not destroy the local self-organization, but rather modulates it so that it becomes *coherent* with the other map organization.

Finally, having this model based on the self-organization of information prototypes leads implicitly to an organization that can be interpreted since it is easy to see what a unit is tuned on. This might be useful for further qualitative comparisons with real fMRI activations.



## 5 Discussion

The model presented in this paper is designed for a general cortically-inspired associative learning. It is based on the cooperation of several self-organizing maps, that are connected one with the other. The design of this model stresses some computational points, that keeps the model functionally close to the biology. The first one is locality, since each unit computes its status from the units it is connected to, without any superior managing process. This leads to the set up of a distributed competition mechanism, whose emerging effect is the rise of a bubble of activity at locally relevant places in the map. The second computational point the model stresses is stripe connectivity between maps. From a strictly computational point of view, this keeps the number of connections under combinatorial explosion. Moreover, the consequent computation has more interesting properties. Using stripes actually constraints the model to overcome partial connectivity by organizing the maps so that related information stands at connected places. This is supported by resonance between cortical layers, and leads to organize states in each map according to the organization of the maps it is connected to. This dependency isn't explicitly given to the model, it can be viewed as a side effect of the shortage of connections. This effect has been observed in our previous work [17] concerning the arm guidance, but it wasn't of primary importance in that context. Last, the novelty here is that our model now uses a Widrow-Hoff learning rule, so that it manages multiple associations between the units of the different maps in the model. Thus, multiple associations between inputs do not anymore require a duplication of these inputs, in terms of different units, on the model's cortical maps.

However, in the present paper, the property of joint organization the model exhibits is reported in the framework of semantic coding observed in cortical areas, since high level word representation appears to be organized according to the body part the word refers to. The ability of our model to generate such kind of organization without any supplementary specification supports its relevance as a functional model of cortical computation, in spite of sometimes less plausible computational mechanisms that keeps the model tractable for a large amount of units.

Considering self-organization of many interconnected self-organizing modules leads to discuss the organization of representations at a global level, that may appear rather more abstract than the organization resulting from the mapping of a mono-modal distribution, as performed by usual unsupervised learning techniques. In the context of autonomous robotics, that this model addresses, anything that is learned is obviously a situated representation. Moreover, the model makes the organization of a particular module in the architecture, dealing with one specific modality, be understandable according to the other modules, and more generally according to the global purpose of such an architecture to address situated behavior. This raises the hypothesis that the influence of the global behavioral purpose at the level of each modality representation is the very property that endows this representation with a semantic value. Therefore, this view of semantics, inspired from biological facts about cortical areas involved in

language, appears to be tractable by a joint organizing model, and to be more generally suitable for any situated multimodal processing in robotics.

## Acknowledgments

The authors wish to thank the Lorraine region, the Robea program of the CNRS and the European MirrorBot project for their contribution.

## References

1. Rolls, E.: *The Brain and Emotion*. Oxford University Press: Oxford (1999)
2. Burnod, Y.: *An adaptive neural network : the cerebral cortex*. Masson (1989)
3. Ballard, D.H.: Cortical connections and parallel processing : Structure and function. *The Behavioral and Brain Sciences* **9** (1986) 67–129
4. Ito, M.: *The cerebellum and neural control*. New-York, Raven (1984)
5. Mountcastle, V.B.: An organizing principle for cerebral function. The unit module and the distributed system. In: *The mindful brain*, Cambridge, MIT Press. (1978)
6. Doya, K.: What are the computations in the cerebellum, the basal ganglia, and the cerebral cortex. *Neural Networks* **12** (1999) 961–974
7. Kohonen, T.: *Self-Organization and Associative Memory*. Springer-Verlag (1988)
8. D.J.Willshaw, von der Malsburg, C.: How parrerned neural connections can be set up by self-organization. In: *Proceedings of the royal society of London*. Volume B 194. (1976) 431–445
9. Kohonen, T., Oja, E.: Visual feature analysis by the self-organising maps. *Neural Computing and Applications* **7** (1998) 273–286
10. Kohonen, T.: The neural phonetic typewriter. *Computer* **21** (1988) 11–22
11. R. Miikkulainen, J.A. Bednar, T.C., Sirosh, J.: Self-organization, plasticity, and low-level visual phenomena in a laterally connected map model of the primary visual cortex. In R.L. Goldstone, P.S., Medin, D., eds.: *Psychology of Learning and Motivation* (36: perceptual learning). San Diego, CA: Academic Press (1997) 257–308
12. Ritter, H., Martinetz, T., Schulten, K.: *Neural Computation and Self-Organizing Maps: An Introduction*. Addison-Wesley Longman Publishing Co. (1992)
13. T. Allard, S.A. Clark, W.J., Merzenich, M.: Reorganization of somatosensory area 3b representations in adult owl monkeys after digital syndactyly. *J Neurophysiol.* **66** (1991) 1048–58
14. Bach-y-Rita, P.: Tactile sensory substitution studies. *Ann NY Acad Sci.* **1013** (2004) 83–91
15. Pulvermüller, F.: *The Neuroscience of Language*. Cambridge University Press (2003)
16. Ménard, O., Frezza-Buet, H.: Rewarded multi-modal neuronal self-organization: Example of the arm reaching movement. In: *Proc. AISTA*. (2004)
17. Ménard, O., Frezza-Buet, H.: Multi-map self-organization for sensorimotor learning: a cortical approach. In: *Proc. IJCNN*. (2003)
18. Amari, S.I.: Dynamical study of formation of cortical maps. *Biological Cybernetics* **27** (1977) 77–87
19. Taylor, J.G.: Neural networks for consciousness. *Neural Netowrks* **10** (1997) 1207–1225

20. Guigon, E., Dorizzi, B., Burnod, Y., Schultz, W.: Neural correlates of learning in the prefrontal cortex of the monkey: A predictive model. *Cerebral Cortex* **5** (1995) 135–147
21. Grossberg, S.: Adaptive pattern classification and universal recoding, i: parallel development and coding of neural feature detectors. *Biological Cybernetics* **23** (1976) 121–134