A Neural Model of Object Naming

Alessio Plebe

Abstract— One astonishing capability of humans is to recognize thousands of different objects visually, and to learn the semantic association between those objects and words referring to them. This work is an attempt to build a computational model of such capacity, simulating the process by which infants learn how to recognize objects and words through exposure to visual stimuli and vocal sounds. One of the main fact shaping the brain of a newborn is that lights and colors come from entities of the world. Gradually the visual system learn which light sensations belong to same entities, despite large changes in appearance. This experience is common between humans and several other mammals, like non-human primates. But humans only can recognize a huge variety of objects, most manufactured by himself, and make use of sounds to identify and categorize them. The aim of this model is to reproduce these processes in a biologically plausible way, by reconstructing the essential hierarchy of cortical circuits on the visual and auditory neural paths.

Keywords— Auditory cortex, object recognition, self-organizing maps

I. INTRODUCTION

THIS work is an attempt to build a model of object naming, and how object recognition and name association can emerge by exposure to an environment reach of objects and vocal sounds.

One of the main fact shaping the brain of a newborn is that lights and colors come from entities of the world. The existence of objects and their individuality is confirmed by other sensorial modalities, like touch, and the possibility to move and interact with some object, but the visual input is the main responsible for building a world representation. Gradually the visual system learn which light sensations belong to same entities, despite large changes in appearance. This experience is common between humans and several other mammals, like non-human primates. But humans only can recognize a huge variety of objects, most manufactured by himself, and make use of sounds to identify and categorize them. Another fact that, not later that the former, begins to shape the brain of newborns is the presence in the environment of special sounds, charged with intentionality. The auditory cortex refine itself so to recognize the most common sound patterns heard, and finally recognize distinct words. This is the time for grasping the most difficult fact: that words sometimes refer to objects.

It is not by chance that the majority of the early vocabulary is made by name of objects, the ones frequently seen by young humans [13].

The interest in this work is to investigate this wonderful phenomena by computational models able to shed light on the main theater of these events: the cerebral cortex. The challenge is to try at the same time to address a very complex

Alessio Plebe is with the Department of Cognitive Science – University of Messina, ITALY, (e-mail: alessio.plebe@unime.it)

high cognitive task, and to adhere as much as possible to the reality of the computations taking place in the brain. Necessarily the level of details will be far from the natural neural networks involved in object cognition and naming. Nevertheless, the biological plausibility is the only road to mitigate the typical risk of all computational explanations: the arbitrariness of the solution. Many different mathematical constructs can approximate the same function, but only one is actually realized in the biological system.

As far as our knowledge this is the first attempt in this direction. There are several neural models of visual object recognition [10], [33], [9]. Not many neural models have been proposed for the auditory process [27], [39], and too little is known yet about the kind of computation leading to word recognition.

There are few computational model of lexical acquisition from auditory and visual stimuli previously proposed [34], but without any relation with the relevant brain processes.

There is a theoretical perspective in this work that could make modeling even less arbitrary and closer to neural realism. It is avoided any design of the mature functions reached by the organism, the emergence of the final function in each component of the system is leaved to the plastic development of the neural circuits. In the cortex, there is very little differentiation in the computational capability that neural circuits will potentially perform in the mature stage. The interaction between environmental stimuli and some basic mechanisms of development is what drives differentiation in computational functions. This position has large empirical support [17], [18], [23], and not only is compatible with the current knowledge about neural genetics [32], but appears to be a correct road for the understanding of the complex interactions between genetic expression and neural plasticity [25].

In pursuing those plans this work had to face with a large gap between the available knowledge of cortical functions in the visual and the auditory paths. Therefore while for the vision subsystem it has been possible to replicate with a certain degree of realism the layout of the main cortical components, for the auditory path the structure of the components and the resulting functions are much more abstract and speculative.

II. THE PROPOSED MODEL

The model will be here explained by first describing the mathematics common to all modules, and then by outlining the overall model, with details of the visual and the auditory paths.

A. The mathematical abstraction of the cortical maps

All the modules composing this model are implemented as artificial cortical maps, adopting the LISSOM (*Laterally*
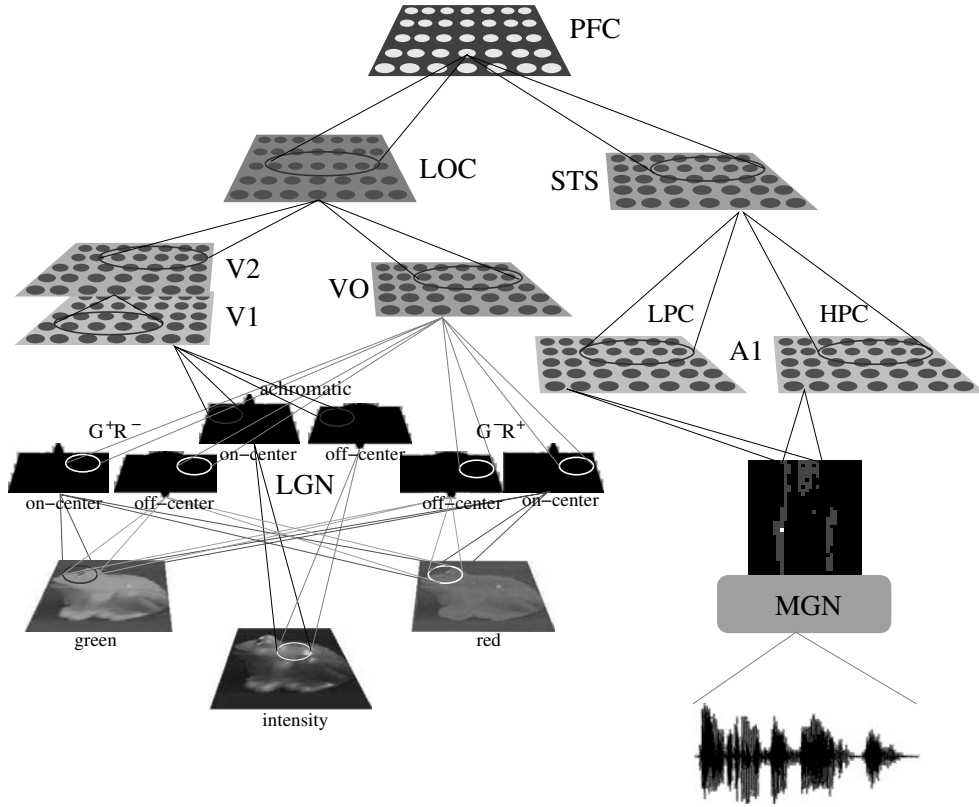
Fig. 1. Overall scheme of the model. All acronyms are listed in Tab. I

Interconnected Synergetically Self-Organizing Map) architecture [36], [3]. This architecture has been chosen because of its reproduction of the neural plasticity, by the combination of Hebb's principle and neural homeostasis, and because is a good compromise between a number of realistic features included, and the simplicity necessary for building complex models. The LISSOM is a two dimensional arrangement of neurons, where each cell is not just connected with the afferent input vector, but receives excitatory and inhibitory inputs from several neighbor neurons on the same map:

$$x_i^{(k)} = f\left(\frac{\gamma_\mathrm{A}}{1 + \gamma_\mathrm{N}\vec{I}\cdot\vec{v}_{r_\mathrm{A},i}}\vec{a}_{r_\mathrm{A},i}\cdot\vec{v}_{r_\mathrm{A},i} + \gamma_\mathrm{E}\vec{e}_{r_\mathrm{E},i}\cdot\vec{x}_{r_\mathrm{E},i}^{(k-1)}\right.$$
$$\left. - \gamma_\mathrm{H}\vec{h}_{r_\mathrm{H},i}\cdot\vec{x}_{r_\mathrm{H},i}^{(k-1)}\right),$$

(1)

where $x_i^{(k)}$ is the activation of the neuron i at time step k. All vectors are composed by a circular neighborhood of given radius around the neuron i: vectors $\vec{x}^{(k-1)}$ are activations of neurons on the same layer at the previous time step. Vector $\vec{v}_{r_\mathrm{A},i}$ comprises all neurons in the underlying layer, in a circular area centered on the projection of i on this layer, with radius r_A. Vectors $\vec{a}_{r_\mathrm{A},i}$, $\vec{e}_{r_\mathrm{E},i}$, and $\vec{h}_{r_\mathrm{H},i}$ are composed by all connections strengths of, respectively afferent, excitatory or inhibitory neurons projecting to i, inside circular areas of radius r_A, r_E, r_H. Vector $\vec{I}$ is just a vector of 1's of the same dimension of $\vec{v}_{r_\mathrm{A},i}$. The scalars γ_A, γ_E, and γ_H, are constants modulating the contribution of afferent, excitatory

and inhibitory connections. The scalar γ_N controls the setting of a push-pull effect in the afferent weights, allowing inhibitory effect without negative weight values. Mathematically, it represents dividing the response from the excitatory weights by the response from a uniform disc of inhibitory weights over the receptive field of neuron i. The map is characterized by the matrices $\mathbf{A}, \mathbf{E}, \mathbf{H}$, which columns are all vectors $\vec{a}$, $\vec{e}$, $\vec{h}$ for every neuron in the map. The function f is a monotonic non-linear function limited between 0 and 1. The final activation value of the neurons is assessed after a certain settling time K.

All connections strengths to a neuron i adapt by following the rules:

$$\Delta\vec{a}_{r_\mathrm{A},i} = \frac{\vec{a}_{r_\mathrm{A},i} + \eta_\mathrm{A}x_i\vec{v}_{r_\mathrm{A},i}}{\|\vec{a}_{r_\mathrm{A},i} + \eta_\mathrm{A}x_i\vec{v}_{r_\mathrm{A},i}\|} - \vec{a}_{r_\mathrm{A},i}, \quad (2)$$

$$\Delta\vec{e}_{r_\mathrm{E},i} = \frac{\vec{e}_{r_\mathrm{E},i} + \eta_\mathrm{E}x_i\vec{x}_{r_\mathrm{E},i}}{\|\vec{a}_{r_\mathrm{E},i} + \eta_\mathrm{E}x_i\vec{x}_{r_\mathrm{E},i}\|} - \vec{e}_{r_\mathrm{E},i}, \quad (3)$$

$$\Delta\vec{h}_{r_\mathrm{H},i} = \frac{\vec{h}_{r_\mathrm{H},i} + \eta_\mathrm{A}x_i\vec{x}_{r_\mathrm{H},i}}{\left\|\vec{h}_{r_\mathrm{H},i} + \eta_\mathrm{A}x_i\vec{x}_{r_\mathrm{H},i}\right\|} - \vec{h}_{r_\mathrm{H},i}, \quad (4)$$

where $\eta_{\{\mathrm{A,E,H}\}}$ are the learning rates for afferent, excitatory and inhibitory synaptic modifications. All rules are based on the Hebb law, with an additional competitive factor, here implemented as a normalization, that maintains constant the integration of all connection strengths to same neuron, and to the same type (afferent, excitatory or inhibitory). This is a computational account for the biological phenomena of

TABLE I
LEGEND AND SIZE OF ALL LAYERS COMPOSING THE MODEL.

LGN	Lateral Geniculated Nucleus	120×120
MGN	Medial Geniculated Nucleus	32×32
V1	Primary Visual Cortex	96×96
V2	Secondary Visual Cortex	30×30
A1	Auditory Primary Cortex	24×24
VO	Ventral Occipital	30×30
LOC	Lateral Occipital Complex	16×16
STS	Superior Temporal Sulcus	16×16
PFC	Pre-Frontal Cortex	12×12

TABLE II
FINAL PARAMETERS USED FOR THE CORTICAL LAYERS.

layer	r_A	r_E	r_H	γ_A	γ_E	γ_H	γ_N
V1	8.5	1.5	7.0	1.5	1.0	1.0	0.0
V2	7.5	8.5	3.5	50.0	3.2	2.5	0.7
VO	24.5	4.0	8.0	1.8	1.0	1.0	0.0
LPC	3.5	2.5	5.5	5.0	5.0	6.7	0.8
HPC	3.5	2.5	5.5	5.0	5.0	6.7	0.8
LOC	6.5	1.5	3.5	1.2	1.0	1.5	0.0
STS	3.5	2.5	2.5	2.0	1.6	2.6	0.0
PFC	3.5	1.5	2.5	1.0	1.0	1.0	0.0

homeostatic plasticity, that induce neurons in the cortex to maintain an average firing rate by correcting their incoming synaptic strengths.

B. The overall model

A sketch of all modules composing the model is visible in Fig. 1. The name of all the components and their dimension is in Tab. I. There are two main paths, one for the visual process and another for the auditory channel. Both paths include thalamic modules, which are not the objective of this study, and therefore are hardwired according to the knowledge of their functions. The two paths joint in the map called PFC. Unlike the lower maps, this is an abstraction of processes actually involving several brain areas in a complex way. Its name is due to evidences of the prefrontal cortex as a locus of multimodal categorical object representations [12], [1].

The functional parameters for the cortical layers are shown in Tab. II. Some of the parameters, like r_E or γ_A, are changed during the training phase, the table presents the final values only.

C. The visual pathway

As visible in Fig. 1, the architecture used here includes hardwired extracortical maps with simple on-center and off-center receptive fields. There are three pairs of sheets in the LGN maps: one connected to the intensity image plane, and the other two connected to the medium and long wavelength planes. The shape of the receptive field for the achromatic channel is given, in a two dimensional coordinates r and c of the retina, as:

$$x^{(r,c)} = -\frac{r^2 + c^2 - \sigma^2}{\sigma^4} e^{-\frac{r^2+c^2}{2\sigma^2}} \qquad (5)$$

In the color channels the internal excitatory portion of the receptive field is connected to the channel of one color, and the surrounding inhibitory part to the opposite color. The cortical process proceeds along two different streams: the achromatic component is connected to the primary visual map V1 followed by V2, the two spectral components are processed by the map VO, the color center, called sometimes also hV4 or V8 [7]. The two streams rejoin in the cortical map LOC, the area recently suggested as the first involved in object recognition in humans [24], [16]. Details of the visual path are in [30], [31].

D. The auditory pathway

The hardwired extracortical MGN component is just a placeholder for the spectrogram representation of the sound pressure waves, which is extracted with tools of the *Festival* software [5]. It is justified by evidences of the spectro-temporal process performed by the cochlear-thalamic circuits [11]. The auditory primary cortex is simulated by a double sheet of neurons, to take into account a double population of cells found in this area [2], where the so-called LPC (*Low-Probability Connections*) is sensitive to the stationary component of the sound signal and the HPC (*High-Probability Connections*) population responds to transient inputs mainly. The next map in the auditory path of the model is STS, because the superior temporal sulcus is believed to be the main brain area responsive to vocal sounds [4]. It should be noted that this is the first use on the LISSOM architecture in non visual cortical maps. To adopt LISSOM in the auditory path is fully justified by evidences of a topological organization in two dimensions of the auditory cortex [22], [21].

III. RESULTS

In this section the training strategies of the experiments will be briefly described, and then the main results achieved will be presented.

A. Exposure to stimuli

The visual path in the model develops in two stages. At the beginning the inputs to the network are synthetic random blobs, simulating pre-natal waves of spontaneous activity, known to be essential in the early development of the visual system [35]. In the second stage, corresponding to the period after eyes opening, natural images are used. In order to address one of the main problem in recognition, the identity of an object under different views, the COIL-100 collection has been used [26] where for each of the 100 objects 72 different views are available.

In the auditory path there are different stages too. At the beginning the maps are exposed to random patches in frequency-time domain, with shorter duration for HPC and longer for LPC. Later all the auditory maps are exposed to the 7200 most common English words (from `http://www.bckelk.uklinux.net/menu.html`) with length between 3 and 10 characters. All words are converted

Fig. 2. Functional organizations emerging in the maps of the model.
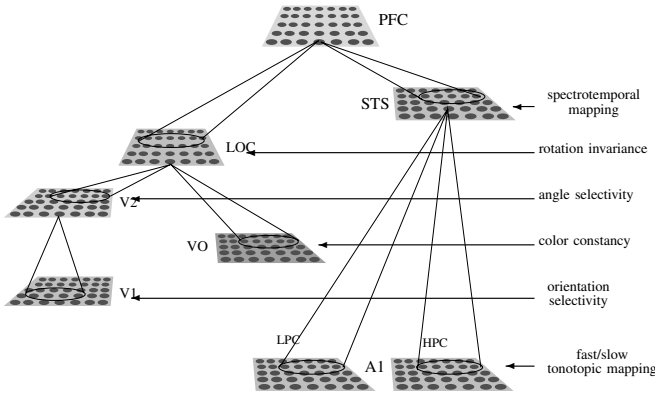
from text to waves using *Festival* [5], with cepstral order 64 and a unified time window of 2.3 seconds.

Eventually the last stage of the training simulate events when simultaneously an object is viewed, and a word corresponding to its basic category is heard. The 100 objects has been grouped manually into 38 categories. Some category, like `cup` or `medicine` count 5 exemplars in the object collection, while others, like `telephone`, have only one exemplar.

B. Functions developed in the cortical maps

At the end of the development each map in the model has evolved its own function, a synthesis is given in Fig. 2. It is remarkable that different functions have emerged from identical computational architectures, the differences are due to:

- different localization of a map in the hierarchy of modules;
- different exposure to environmental stimuli;
- different structural parameters, as from Table II.

Now all the functions obtained in the experiment will be discussed, with reference to Fig. 2, starting with the visual path.

Orientation selectivity is the main organization in the primary visual cortex, where the responsiveness of neurons to oriented segments is arranged over repeated patterns of gradually changing orientations, broken by few discontinuities [6], [37]. This sort of arrangement emerge in the model V1, as already demonstrated in [36] and [31]. In the secondary visual cortex the main phenomena recently discovered [15] is the selectivity to angles, especially in the range between 60 and 150 degrees. This kind of selectivity is acquired in the model map V2, details are in [29]. Color constancy is is the tendency of the color of a surface to appear more constant than it is in reality. This property is helpful in object recognition, and develops sometimes between two and four months of age [8]. In the model VO only the essential feature of color constancy is reproduced, which is the ability of neurons to respond to specific hues, regardless of the intensity. One of the main function shown by the layer LOC in the model is visual invariance, the property of neurons to responding to peculiar

object features despite changes in the object's appearance due to different points of view. Invariance indeed is one of the main requirement for an object-recognition area, and is found in the human LOC [14], [16]. More details on the invariance properties in the model LOC are in [28].

Tonotopic mapping is a known feature of the primary auditory cortex to represent the dimensions of frequency and time sequence in a sound pattern [38]. In the model is split into a sheet where neurons have receptive fields more elongated along the time dimension (LPC) and another where the resulting receptive fields are more elongated along the frequency dimension (HPC). The spectrotemporal mapping obtained in STS is a population coding of features, in frequency and time domains, representative of the sound patterns heard during the development phase. Therefore it reflects the statistical phonemic regularities in the common spoken English, extracted from the 7200 samples used in the training. As mentioned in the introduction, it is difficult to include the reproduction of more specific and detailed cortical functions, operating in phoneme and syllable categorization, since almost nothing is known about such functions in the brain yet.

C. Recognition and categorization in PFC

The upper map PFC in the model reflects how the system has learned the association of certain sound forms with the visual appearances of objects. In order to evaluate what has been achieved in the PFC map its content has been analyzed by clustering into a ordinary SOM (*Self Organized Map*) [19]. Being o an object of the COIL set $\mathcal{O}$, $\mathcal{W}$ the set of names of categories, and x a node of the SOM, several labeling functions can be established:

$$l^{(I|c)}(x) = \arg\max_{o \in \mathcal{O}} \left\{ \left| \left\{ I_i^{(o)} : x = v\left(I_i^{(o)}, c(o)\right) \right\} \right| \right\}, \tag{6}$$

$$l^{(c|I)}(x) = \arg\max_{c \in \mathcal{W}} \left\{ \left| \left\{ I_i^{(o)} : c = c(o) \wedge x = v\left(I_i^{(o)}, c\right) \right\} \right| \right\}, \tag{7}$$

$$l^{(I|u)}(x) = \arg\max_{o \in \mathcal{O}} \left\{ \left| \left\{ I_i^{(o)} : u \neq c(o) \wedge x = v\left(I_i^{(o)}, u\right) \right\} \right| \right\}, \tag{8}$$

$$l^{(I)}(x) = \arg\max_{o \in \mathcal{O}} \left\{ \left| \left\{ I_i^{(o)} : x = v\left(I_i^{(o)}, \epsilon\right) \right\} \right| \right\}, \tag{9}$$

$$l^{(c)}(x) = \arg\max_{c \in \mathcal{W}} \left\{ \left| \left\{ c : x = v\left(\epsilon, c\right) \right\} \right| \right\}, \tag{10}$$

with $I_i^{(o)}$ an image of the COIL database representing object o at viewpoint i, $c(o) : \mathcal{O} \to \mathcal{W}$ the lexical category of the object o, and $v(\cdot, \cdot)$ the function associating an image and a word given as input to the model with a winner neuron in the SOM. The input ϵ to $v(\cdot, \cdot)$ means null input. The notation $|\cdot|$ is used here as the cardinality of a set. The labeling of the joint recognition of objects by visual aspect and category is $l^{(I,c)}(\cdot)$ from (6), the labeling of recognized categories of objects is $l^{(c,I)}(\cdot)$ from (7).
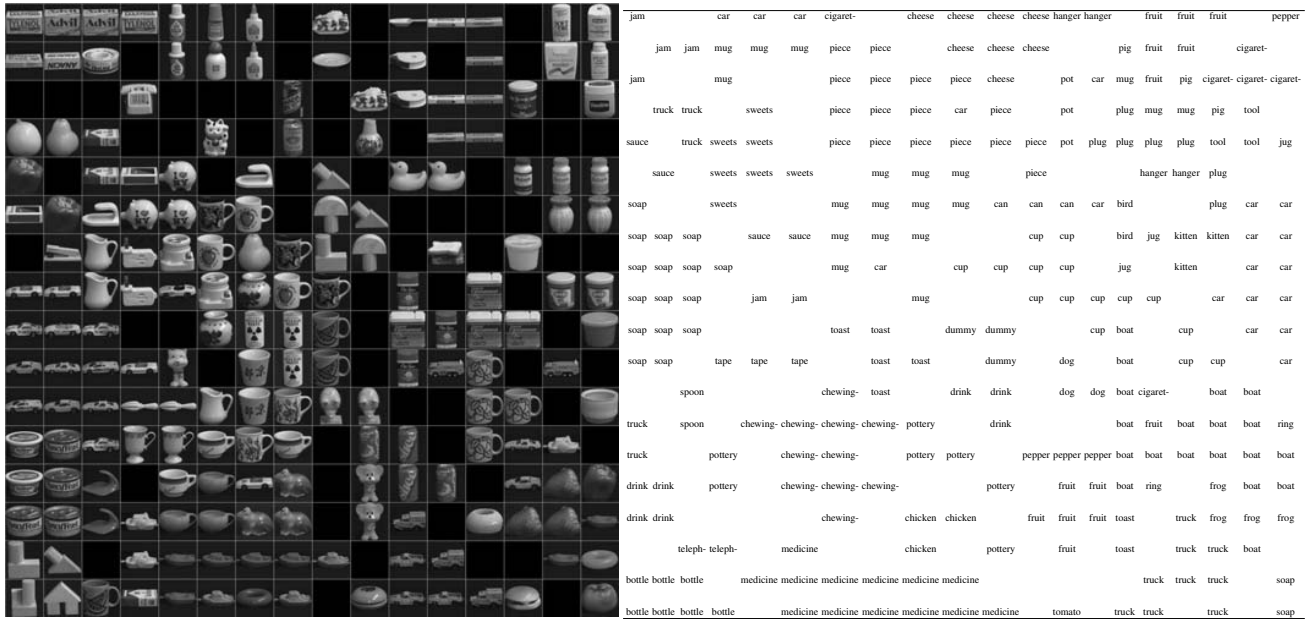
Fig. 3.   Organization of visual (left) and lexical (right) categories from PFC, as revealed by SOM clustering.

TABLE III

ACCURACY IN RECOGNITION ACHIEVED BY PFC, FOR THE RECOGNITION CONDITIONS SEE THE REFERENCED EQUATIONS IN THE TEXT.

rcognition conditions	equation	accuracy
vision + word	(6)	0.88
categories	(7)	0.98
vision - word	(8)	0.07
images only	(9)	0.16
words only	(10)	0.55

These two labeling are shown in Fig. 3, and give the idea of the categorization within the visual and the lexical space. In both maps there is a strong evidence of clustering of similar objects, but in the visual map there are more cases when group of similar objects are split into more then one cluster. It is the case of the wooded pieces visible in the middle and also in the lower right corner. This is a typical phenomena of objects with strong dissimilarity between two or more ranges of viewpoints. There are less split clusters of categories in the lexical map.

From the labeling functions immediately follows the possibility of estimate the accuracy of recognition, simply by weighting the number of cases where the category or the object has been classified as the prevailing one in each node of the SOM. For example, in the case (6), recognition of objects by visual aspect and category, the corresponding accuracy is:

$$a^{(I,c)}(o) = \frac{\left| \left\{ I_i^{(o)} : l\left(v\left(I_i^{(o)}, c(o)\right)\right) = o \right\} \right|}{\left| \left\{ I_i^{(o)} \right\} \right|}. \qquad (11)$$

The accuracy of the various labeling functions are shown in Table III. The first two rows clearly prove that the system

has learned a high capacity of object recognition and naming, with respect to the small world of object and names of the experiment. The third line is the accuracy for the case where a word is heard while an object of different category is seen. It is evident how the mismatch of the name almost hampers the correct recognition. It has to be taken into account that for the COIL library, the accuracy of recognition by chance is of 0.01, therefore 0.07 is still higher that by chance, but drastically reduced from the case of joint recognition. The last two lines are the cases when one of the two inputs is missed. Being the outcome of PFC, these accuracy figures are not the accuracy of the separate outcome of the visual or the auditory paths, which would be higher. Those figures can be interpreted as the chance of guessing a category name by seeing a related object, or of imaging an object view by hearing its name. The larger accuracy in guessing a correct object from the category name is also a consequence of the small number of categories, 38, compared with object views, 7200.

IV. CONCLUSIONS

The model here described is a first attempt in simulating the interaction of the visual and the auditory cortex in learning object recognition and naming. Being a model of high level complex cognitive functions, it necessarily lacks several details of the biologic cortical circuits. It lacks even more biological plausibility in the auditory path because of the current knowledge of the processes going on there. Moreover, being a first attempt, it is certainly oversimplified in aspects that can be improved and refined in further development of this model. Examples can be the inclusion of backprojections between maps in the hierarchy, and trials on preliminary categorization at the level of phonemes and syllable in the auditory path. An important result achieved so far is the emergence of naming and recognition abilities only by exposing the system to environmental simulations, in term of pre-natal spontaneous activities, and later to natural images and vocal sounds. For

this reason the model is believed to be a useful computational tool for future investigations on phenomena known in developmental psychology, like the effect of shape in learning object names [20], [13].

REFERENCES

[1] I. Akirav and M. Maroun. Ventromedial prefrontal cortex is obligatory for consolidation and reconsolidation of object recognition memory. *Cerebral Cortex*, 16:1759–1765, 2006.

[2] M. Atzori, S. Lei, D. I. P. Evans, P. O. Kanold, E. Phillips-Tansey, O. McIntyre, and C. J. McBain. Differential synaptic processing separates stationary from transient inputs to the auditory cortex. *Neural Networks*, 4:1230–1237, 2001.

[3] J. A. Bednar. *Learning to See: Genetic and Environmental Influences on Visual Development*. PhD thesis, University of Texas at Austin, 2002. Tech Report AI-TR-02-294.

[4] P. Belin, R. J. Zatorre, and P. Ahad. Human temporal-lobe response to vocal sounds. *Cognitive Brain Research*, 13:17–26, 2002.

[5] A. W. Black and P. A. Taylor. The festival speech synthesis system: System documentation. Technical Report HCRC/TR-83, Human Communciation Research Centre, University of Edinburgh, Edinburgh, UK, 1997.

[6] G. G. Blasdel. Orientation selectivity, preference, and continuity in monkey striate cortex. *Journal of Neuroscience*, 12:3139–3161, 1992.

[7] A. A. Brewer, J. Liu, A. R. Wade, and B. A. Wandell. Visual field maps and stimulus selectivity in human ventral occipital cortex. *Nature Neuroscience*, 8:1102–1109, 2005.

[8] J. L. Dannemiller. A test of color constancy in 9- and 20-weeks-old human infants following simulated illuminant changes. *Developmental Psychology*, 25:171–184, 1989.

[9] G. Deco and E. Rolls. A neurodynamical cortical model of visual attention and invariant object recognition. *Vision Research*, 44:621–642, 2004.

[10] S. Edelman and S. Duvdevani-Bar. A model of visual recognition and categorization. *Philosophical transactions of the Royal Society of London*, 352:1191–1202, 1997.

[11] M. A. Escabi and H. L. Read. Representation of spectrotemporal sound information in the ascending auditory pathway. *Biological Cybernetics*, 89:350–362, 2003.

[12] D. J. Freedman, M. Riesenhuber, T. Poggio, and E. K. Miller. Categorical representation of visual stimuli in the primate prefrontal cortex. *Science*, 291:312–316, 2001.

[13] L. Gershkoff-Stowe and L. B. Smith. Shape and the first hundred nouns. *Child Development*, 75:1098–1114, 2004.

[14] K. Grill-Spector, Z. Kourtzi, and N. Kanwisher. The lateral occipital complex and its role in object recognition. *Vision Research*, 41:1409–1422, 2001.

[15] M. Ito and H. Komatsu. Representation of angles embedded within contour stimuli in area V2 of macaque monkeys. *Journal of Neuroscience*, 24:3313–3324, 2004.

[16] N. Kanwisher. The ventral visual object pathway in humans: Evidence from fMRI. In L. Chalupa and J. Werner, editors, *The Visual Neurosciences*. MIT Press, Cambridge (MA), 2003.

[17] L. C. Katz and E. M. Callaway. Development of local circuits in mammalian visual cortex. *Science*, 255:209–212, 1992.

[18] A. Kirkwood and M. F. Bear. Hebbian synapses in visual cortex. *Journal of Neuroscience*, 14:1634–1645, 1994.

[19] T. Kohonen. *Self-Organizing Maps*. Springer-Verlag, Berlin, 1995.

[20] B. Landau, L. B. Smith, and S. Jones. Syntactic context and the shape bias in children's and adults' lexical learning. *Journal of Memory and Language*, 31:807–825, 1992.

[21] D. R. Langers, W. H. Bacjes, and P. van Dijk. Representation of lateralization and tonotopy in primary versus secondary human auditory cortex. *NeuroImage*, 34:264–273, 2007.

[22] J. F. Linden and C. E. Schreiner. Columnar transformations in auditory cortex? a comparison to visual and somatosensory cortices. *Cerebral Cortex*, 13:83–89, 2006.

[23] S. Löwel and W. Singer. Experience-dependent plasticity of intracortical connections. In M. Fahle and T. Poggio, editors, *Perceptual Learning*. MIT Press, Cambridge (MA), 2002.

[24] R. Malach, J. B. Reppas, R. R. Benson, K. K. Kwong, H. Jiang, W. A. Kennedy, P. J. Ledden, T. J. Brady, B. R. Rosen, and R. B. Tootell. Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proceedings of the Natural Academy of Science USA*, 92:8135–8139, 1995.

[25] M. Mjdan and C. J. Shatz. Effects of visual experience on activity-dependent gene regulation in cortex. *Neural Networks*, 9:650–659, 2006.

[26] H. Murase and S. Nayar. Visual learning and recognition of 3-d object by appearence. *International Journal of Computer Vision*, 14:5–24, 1995.

[27] C. Näger, J. Storck, and G. Deco. Speech recognition with spiking neurons and dynamic synapses: a model motivated by the human auditory pathway. *Neurocomputing*, 44–46:937–942, 2002.

[28] A. Plebe. Learning visual invariance. In M. Verleysen, editor, *ESANN 2006 – 14th European Symposium on Artificial Neural Networks*, pages 71–76, Evere (BE), 2006. d-side Publications.

[29] A. Plebe. A model of angle selectivity development in visual area v2. *Neurocomputing*, in press.

[30] A. Plebe and R. G. Domenella. The emergence of visual object recognition. In W. Duch, J. Kacprzyk, E. Oja, and S. Zadrony, editors, *Artificial Neural Networks – ICANN 2005 15th International Conference, Warsaw*, pages 507–512, Berlin, 2005. Springer-Verlag.

[31] A. Plebe and R. G. Domenella. Early development of visual recognition. *BioSystems*, 86:63–74, 2006.

[32] S. R. Quartz. Innateness and the brain. *Biology and Philosophy*, 18:13–40, 2003.

[33] M. Riesenhuber and T. Poggio. Models of object recognition. *Nature Neuroscience*, 3:1199–1204, 2000.

[34] D. Roy and A. Pentland. Learning words from sights and sounds: a computational model. *Cognitive Science*, 26:113–146, 2002.

[35] F. Sengpiel and P. C. Kind. The role of activity in development of the visual system. *Current Biology*, 12:818–826, 2002.

[36] J. Sirosh and R. Miikkulainen. Topographic receptive fields and patterned lateral interaction in a self-organizing model of the primary visual cortex. *Neural Computation*, 9:577–594, 1997.

[37] W. Vanduffel, R. B. Tootell, A. A. Schoups, and G. A. Orban. The organization of orientation selectivity throughout the macaque visual cortex. *Cerebral Cortex*, 12:647–662, 2002.

[38] C. Verkindt, O. Bertrand, F. Echallier, and J. Pernier. Tonotopic organization of the human auditory cortex: N100 topography and multiple dipole model analysis. *Electroencephalography and Clinical Neurophisiology*, 96:143–156, 1995.

[39] M. Volkmer. A pulsed neural network model of spectro-temporal receptive fileds and population coding in auditory cortex. *Neural Computing*, 3:177–193, 2004.