



HAL
open science

A three-layered cortical network encodes identity and abstract categorical structure of behavioral sequences as in the primate lateral prefrontal cortex

Xavier Hinaut, Peter Dominey

► To cite this version:

Xavier Hinaut, Peter Dominey. A three-layered cortical network encodes identity and abstract categorical structure of behavioral sequences as in the primate lateral prefrontal cortex. Cinquième conférence plénière française de Neurosciences Computationnelles, "Neurocomp'10", Aug 2010, Lyon, France. hal-00553435

HAL Id: hal-00553435

<https://hal.science/hal-00553435>

Submitted on 16 Mar 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A Three-Layered Cortical Network Encodes Identity and Abstract Categorical Structure of Behavioral Sequences as in the Primate Lateral Prefrontal Cortex

Xavier Hinaut, Peter Ford Dominey
Stem Cell and Brain Research Institute, INSERM U846
18 ave Doyen Jean Lepine, 69675 Bron Cedex
France
{xavier.hinaut,peter.dominey}@inserm.fr

ABSTRACT

Categorical encoding is crucial for mastering large bodies of related sensory experiences. Recent single-unit recording studies in the macaque prefrontal cortex have demonstrated two characteristic forms of neural encoding of the sequential structure of the animal's behaviour. One population of neurons encodes the specific behavioural sequences. A second population of neurons encodes the sequence category (e.g. ABAB, AABB or AAAA) and does not differentiate sequences within the category [1]. Interestingly these neurons are intermingled in the lateral prefrontal cortex, and not topographically segregated. Here we report on a neural network simulation study that reproduces and explains these results. We simulate a cortical circuit as three 5x5 layers (infra-granular, granular, and supra-granular) of leaky integrator neurons with a sigmoidal output function, and we examine 10^3 such circuits running in parallel. The model is presented with 11 4-element sequences following Shima et al. We isolated one subpopulation of neurons each of whose activity predicts individual sequences, and a second population that predicts category independent of the specific sequence. We argue that a richly interconnected cortical circuit is capable of internally generating a neural representation of category membership, thus significantly extending the scope of recurrent network computation.

KEY WORDS

Liquid state machine, reservoir computing, temporal recurrent network, laminar structure, sequence learning, primate electrophysiology.

1. Introduction

In 1989 the first report of prefrontal neuronal activity encoding the sequential properties of sensorimotor sequences was reported by Barone and Joseph [2]. They demonstrated that after the macaque had been shown a spatial sequence of three illuminated push-buttons, and was preparing to reproduce the sequence, neurons in the prefrontal cortex encoded both the spatial location of the targets, and most strikingly, their order or rank in the sequence. Thus, the prefrontal neural population was considered to encode the entire

action plan, necessary for the animal to reproduce the visualized sequence. Since then a number of studies have been performed, looking more closely at the sequence specific encoding of visuo-spatial behavioral components of sensorimotor sequences in the prefrontal cortex [e.g. 3-4].

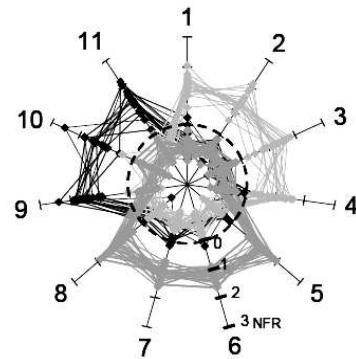


Figure 1. Normalized LPFC neuron activity for three populations of neurons that are selectively active for the three categories of sequences - Paired [1-4], Alternate [5-8] and repeat [9-11] (from [1]).

Recently, Shima and colleagues [1] have considered that when the subject is required to encode a large number of sensorimotor sequences, the increasing memory load would potentially require the animal to adopt a categorization strategy, so the similar sequences would be encoded as members of the same category. They employed a sequence reproduction task, where sequences were made up of the motor actions push, pull and turn, using a manipulable lever. The animal was required to be able to observe and reproduce 11 sequences from three categories: AABB, ABAB and AAAA, where A and B were each systematically replaced by one of the three elements push, pull and turn. 4 sequences were of the category AABB (e.g. Turn, Turn, Push, Push), 4 of the category ABAB and three of the category AAAA. In the experiment, the animal was presented one of these sequences under visual guidance 5 times, and then was required to reproduce the sequence from memory. This was repeated for each of the 11 sequences. Single unit recordings were performed in the lateral prefrontal cortex during the delay preceding the first reproduction of the sequence from memory.

Shima et al [1] found a population of neurons that had the traditional coding of spatio-temporal characteristics of individual sequences. Most interestingly, they found a second population of neurons that were active for one of the three categories, without distinguishing the individual sequences within that category (illustrated in Figure 1). They observed that this categorical information was not directly present in the visual input to the animal. Instead, the brain had somehow extracted the appropriate regularities in order to recognize the repetitive structure that defined the three categories of sequences.

They then raised the crucial question: what are the underlying neuronal processes that would allow these PFC neurons to extract the required properties to form these categorical representations? They propose that the answer to this question will likely be found through the development of neural network models of the PFC applied to this problem.

We have previously developed a neural network model of prefrontal cortex [5,6] with the objective of explaining the electrophysiological findings of the identity and sequential order of sequence elements as initially observed by Barone and Joseph [2]. The principal characteristics of this model were (1) PFC was modeled as two 2-D layers of leaky integrator sigmoidal average firing rate neurons, that were connected by recurrent connections that were randomly selected over the interval $[-0.45, 0.55]$. This mixture of excitatory and inhibitory connections yielded a dynamical system that was sensitive both to the spatial location of targets on the input array, and the previous sequential context. (2) Importantly, the recurrent connections were not subject to modification by learning, instead, learning was used to link these dynamic patterns of activity with the required output states. This was the first expression of the concept that came to be known as the liquid state machine [7], and the subsequent echo state machine [8] and the concept of reservoir computing [9].

Indeed, this dynamical system thus was able to reproduce the behaviour and neurophysiology [5] reported in Barone and Joseph [2]. It also displayed a rich capability to encode and reproduce the serial order and temporal structure of behavioral sequences in a variety of conditions [10-12]. When exposed to sequences with abstract structure such as ABCBAC the model failed to characterize this repetitive structure, leading us to propose an additional working memory mechanism which allowed such sequences to be recoded as an abstract structure $\{x \ x \ x \ n-2 \ n-4 \ n-3\}$ [13].

In the current research our goal is to re-examine the capabilities of a recurrent network to perform categorization as observed by Shima et al [1].

2. Recurrent Network Model

Primate cortex is characterized by its laminar structure, and the connectivity patterns within and between layers [14]. We thus model cortex as a three layered structure of leaky integrator, sigmoid output, average firing rate neurons, illustrated in Figure 2.

Each layer (supra-granular, granular and infra-granular) is topographically organized as a 5×5 sheet.

Based on the structure of divergence of receptive fields across layers, the connections between infra-granular and granular layers are topographic (point to point), while connections between granular and supra-granular (W^{GS}), and between supra-granular and infra-granular (W^{SI}) are divergent. We studied two mechanisms for implementing this divergence. The “matrix” method used a connection matrix that defined unique connections from each neuron in the source layer to all neurons in the target layer. In this method, connections between granular and supra-granular, and between supra-granular and infra-granular are randomly selected from the interval $[-0.5, +0.5]$ with a uniform distribution, and topographic connections set to zero.

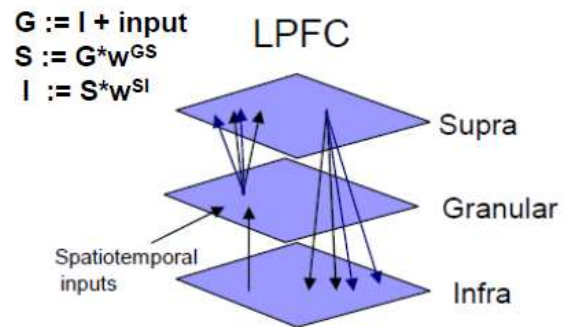


Figure 2. Three-layered cortical model. G, S and I are Granular, Supragranular and Infragranular respectively. Each layer corresponds to a 5×5 matrix of leaky integrator neurons, with the indicated connectivity. The “:=” notation is a compact representation where the right hand side is a differential equation solved by the Euler method, with a 5 ms time step and 100 ms time constant. “*” represents matrix multiplication or convolution. The left hand side corresponds to the membrane potential. The resulting value is passed through a sigmoid function to generate the output average firing rate.

In the second “mask” connectivity scheme that we studied, the projection from each neuron in the source layer to the target layer was defined by a single recurring pattern that was convolved with the source neuron activity and projected to the target layer. Again, values for the mask defining connections between granular and supra-granular, and between supra-granular and infra-granular are randomly selected from the interval $[-0.5, +0.5]$ with a uniform distribution, and topographic connections set to zero. Connections between infra-granular and granular are topographic (point to point). Results will be presented for both the matrix and the mask connectivity models.

Input sequences are presented to the granular layer, in the form of activation of a neuron in the granular layer for 1 second (200 simulation time steps, i.e. STS), followed by a 1 second pause, for each of the four sequence elements. One second (200 STS) after the final target presentation, we consider that the network is in the same state as the primates in Shima [1] prior to execution of the sequence. That is, the sequence has been presented, and the primate is preparing to execute.

We sample the neurons from the three cortical layers at this time, and analyse their ability to identify and categorize the sequences.

3. Model Performance & Results

We now consider the performance of the model in recognizing individual sequences, and in categorizing the sequences. Simulation results are presented for 1000 instances of the matrix and mask models. Each instance has a unique set of randomly assigned connections as described above.

3.1 Sequence Recognition

For each of the 11 sequences, after the sequence has been presented the activity of the 25 neurons in each of the three cortical layers is recorded. We then calculate the sequence discrimination capability of each of the three layers.

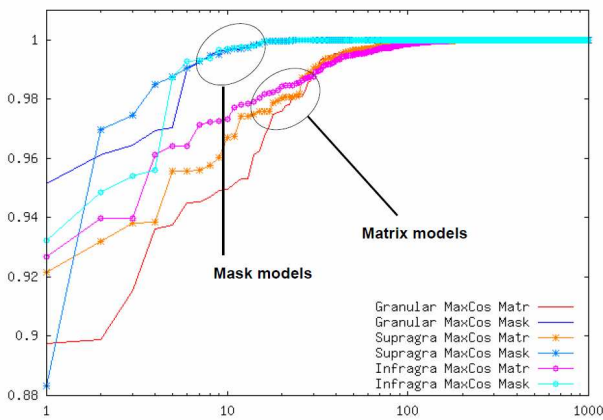


Figure 3. Sequence recognition performance for 1000 model simulations with Matrix and Mask connectivity. Log-linear plot.

For a given layer, we collect the 11 activity vectors generated by each of the sequences. We then calculate the vector cosine of the angle between each pair of vectors in this 11-D space. A cosine of 0 indicates that the vectors are orthogonal, and this corresponds to a high discrimination between the vectors. A cosine of 1 indicates that the vectors are parallel, and corresponds to non-discrimination.

Max(cos) < 0.999	Matrix	Mask
Granular	100	14
Supragranular	106	15
Infra	111	15
Max(cos) < 0.99		
Granular	31	5
Supragranular	29	5
Infra	33	5
Max(cos) < 0.95		
Granular	10	0
Supragranular	4	1
Infra	3	2

Table 1: Performance details for 1000 mask vs. matrix networks discriminating 11 sequences. Values indicate

quantity of models per 1000 with corresponding discrimination.

Figure 3 illustrates the performance for 1000 simulations of the model, with performance decreasing along the X-axis from left to right. We see that a small but significant number of models are capable of discriminating the 11 sequences with a maximum cosine of <0.99 . Indeed we can consider that these relatively small networks ($3 \times 25 = 75$ neurons) can be thought of as running in parallel, thus their discrimination capabilities can be pooled into a larger set of 75K neurons with an additive discrimination capability.

3.2 Sequence Categorization

Shima et al. recorded individual LPFC neurons and characterized their responses to sequences within the three different categories. In order to compare our simulated neurons with theirs, we sampled all of the neurons in the three layers of a run of the 1000 cortical circuits, for a total of 75,000 neurons. We developed a heuristic to isolate neurons that had a strong response to sequences within a category, and weak response to sequences in the other two categories.

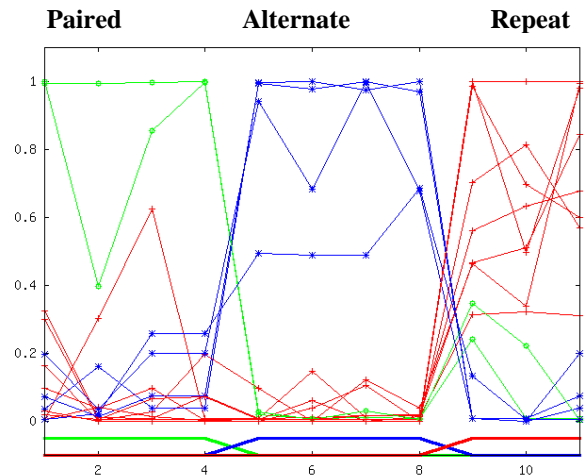


Figure 4. Neural responses to the 11 sequences for the matrix model. Linear version of Fig 1. Example simulated neurons that each respond preferentially to one of the three sequence categories.

Matrix	Paired	Alternate	Repeat	Total
Supra	17	34	60	111
Gran	14	41	39	94
Infra	22	42	64	128
Total	53	117	163	333
Mask				
Supra	12	9	26	47
Gran	9	13	18	40
Infra	17	14	20	51
Total	38	36	64	138

Table 2: Performance details for the 1000 mask vs matrix models categorizing the 11 Sequences as Paired, Alternate and Repeat. Number of neurons per model type and layer that discriminate the given category.

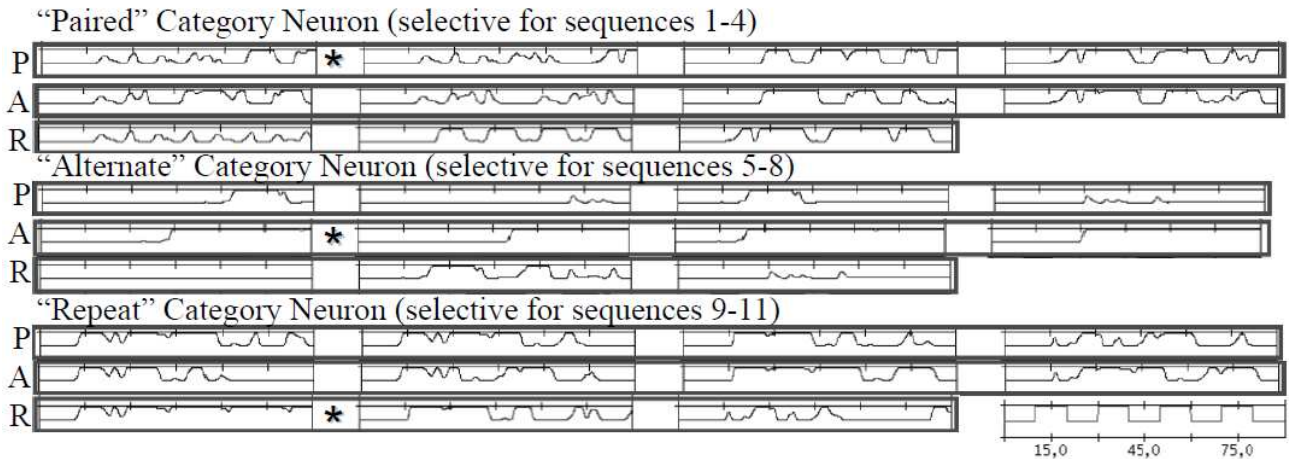


Figure 5. Responses of three neurons with category specific activity. Three sets of lines correspond to activity of three distinct neurons. Each separate line corresponds to responses to sequences from one of the three categories (P, A, R). For each of the three displayed neurons, its preferred category responses are on the line with “*”. Each separate profile on a line corresponds to the neural activity during the presentation of the sequence and the following delay. Lowest right corner, timescale of sequence element presentation (high signal indicates sequence element being presented).

Table 2 illustrates the number of neurons that displayed the desired sequence categorization characteristics. These neurons were identified by their characteristic firing rate during the delay after the sequence presentation such that $\min(X) > \max(Y) \geq \max(Z)$, where X, Y and Z are mapped onto the three categories paired, alternate and repeat. $\min(X)$ refers to the minimum firing rate in response for all of the sequences in that category. Thus, we isolate neurons whose minimal response to the three or 4 sequences in one category are superior to that same neuron’s maximal response to any of the neurons in the remaining two categories. It can be seen that according to this criteria, a total of 333 neurons in the Matrix model and 138 neurons in the mask model perform the sequence categorization.

Figure 4 illustrates this neural activity for an illustrative set of neurons that are each selectively active for one of the three categories. This is a “linear” version of Figure 1 above, from Shima et al. [1]. Neural activity in green (left) illustrates selectivity for the paired (AABB) sequences, blue (middle) for the alternate (ABAB) and red (right) for the repeat (AAAA) sequences. To further examine this category selective activity, we plotted the neural activity for three neurons that were identified to be selective to one of the three respective categories. This neural activity is illustrated in Figure 5.

The “paired” category neuron demonstrates a sustained activation at the end of all paired sequences (sequences on the line marked with “*”). During the same time period, this neuron displays no activation for the “alternate” and “repeat” sequences. The same form of category specific activity is seen for the “alternate” and “repeat” category neurons displayed in Fig 5.

3.3 Effects of Noise

Biological nervous systems are typically not quiescent in the absence of external stimuli, and we thus studied the effects of noise injected into the membrane potential

of each neuron of the three layers of the network on discrimination and categorization in two cases. We examined the effects of noise with a mean value of 0, uniformly distributed over the interval $[-1, 1]$, with gains of 5% and 25%.

For both discrimination and categorization, as the gain on the noise increases, the effect of noise begins to dominate the effects of the signal. Indeed, in the limit case, this can give false positive performance both for discrimination and categorization. That is, the system responds to distinct patterns of noise, rather than to the input sequences.

To control for this, we generated 100 distinct networks (i.e. distinct random connectivity patterns), and presented the 11 sequences over 10 repetitions, and then analysed the neural responses over these 10 repetitions.

For the 25% neutral noise condition, 51 neurons were identified – using a similar heuristic as the one used for Figure 4 – that categorized the sequences across the 100 simulations, each having 10 repetitions. One of these neurons displayed the same category response over two repetitions of the 11 sequences. Importantly, only one of these neurons mistakenly responded to two different categories for two repetitions of the 11 sequences. For the 5% noise condition, 24 neurons had category-specific responses, and none responded to more than one category. Interestingly, we make the paradoxical observation that with increasing noise, more neurons display category-specific responses.

Thus, the addition of noise increases the variability of the responses of these neurons. However, in the vast majority of our observations (100% in the 5% noise condition), category-related neurons do not respond to multiple categories. That is, if a given category-related neuron is active, it identifies the category of the input sequence, even in the presence of noise. Thus, a population of such neurons, with responses distributed over the three categories, would allow a system to accurately categorize input sequences in the presence of noise.

4. Discussion

Neocortex provides an adaptive capability for controlling the “reptilian” basal ganglia and brainstem structures. Cortex cannot anticipate all possible configurations of the perceptual world that can face the organism. Instead, it must embody general coding strategies that can in principal become sensitive to any significant regularities that can arise in the perceptual surfaces. In this context it has been demonstrated exhaustively that recurrent networks of dynamic neuron-like computing elements demonstrate sensitivity to arbitrary sequential structure in the input [5-9].

Interestingly these previous studies have not demonstrated the capability of such recurrent systems to perform the categorical representation of sequential structure as observed in the primate lateral PFC by Shima *et al.* and in our 3-layered cortical network simulation. The exceptions, in which abstract categorical structure can be represented, have required additional representational structure such as working memory in order to detect the abstract structure [e.g. 13, 16]. It is thus of particular interest to determine if in a more general layered architecture, the required properties are inherent without a requirement for specific task-related coding mechanisms for extracting categorical structure. As noted by Shima [1], such categorical information is not directly present in the input, and cannot be directly anticipated, but rather must be generated as a form of meta-representation within the cortex.

Such a representational schema would progress from action to action sequence to action sequence category. In principal, the possible categories are arbitrary and uncountable – that is, in addition to the paired, alternate and repeat categories, one could consider the categories such paired triplets and alternating triplets (e.g. AAABBB, ABABAB). Thus, the space of possible categories is unbounded, and the cortical network must be of such a nature as to encompass this unbounded possibility.

In this context, recurrent connections provide this unbounded re-combinatorial capability, so that a potentially unbounded space of possible categorical structure is present within the system. The current research demonstrates for the first time how such categorical information can be encoded in the activity of single neurons within a structured recurrent neural network.

It is highly likely that the laminar organization of cortex is not arbitrary, but rather represents an evolution of encoding capacity. Indeed, comparative neuroanatomy confirms the progressive laminar complexity of cortex in primate evolution [15]. We thus hypothesize that the current capability to demonstrate categorical encoding in single units within a three-layered cortical network is in part the result of the use of a multiple-layered recurrent system. Future research should make direct comparisons between architectures with different laminar structure including simpler 2- and 1-layered systems, as well as more complex multiple layered systems. Independent of these future results, the current research demonstrates

that within an anatomically motivated three-layered recurrent network, distinct subsets of neurons encode the identity of distinct behavioral sequences, while a separate and distinct population of neurons within the same network encodes the categorical structure of these same sequences independent of their identity. This dual encoding of sequential and abstract structure within the same network helps to provide insight on the underlying neurophysiological mechanisms as suggested by Tanji *et al.* [17].

In this context, a question concerns the ability of the primate and our model to generalize in this task. Further experiments with primates would permit one to demonstrate a possible generalization of sequence categorization to sequences that have not been presented, but nevertheless belong to one of the categories used by Shima *et al.* For example, sequences “Push-Push-Pull-Pull” and “Pull-Pull-Push-Push” which are in the category “paired” have not been used by Shima *et al.* It would be interesting to determine if the neurons which were active for the four sequences of category “paired” would also be firing actively for these two sequences: it would show a generalization of categorization. Generalization with other “inputs” (other movements for instance) could also be tested, thus showing an even higher level of generalization. Further simulations would be then required to determine if our model could also be capable of this hypothetical generalization.

Simulations with noise demonstrate that neurons within the model continue to retain category-specific responses, but with increased variability. Interestingly, up to a limit, increasing noise yields an increase in the number of neurons with category-specific responses. This appears to correspond to situation of stochastic resonance, when increases in levels of random noise cause an increase, rather than a decrease, in a metric of the quality of signal transmission or detection performance [18]. It remains to explore whether our current model can serve as a demonstration that perceptual discrimination and categorization may rely on biological noise.

Acknowledgements

This work is supported by the FP7 Grant 231267 Organic.

References

- [1] Shima, K., Isoda, M., Mushiake, H., & Tanji, J. (2007). Categorization of behavioural sequences in the prefrontal cortex. *Nature*, 445, 315-318.
- [2] Barone P, Joseph J-P (1989) Prefrontal cortex and spatial sequencing in the macaque monkey. *Experimental Brain Research*, 78, 447-464
- [3] Funahashi, S., Inoue, M., & Kubota, K. (1997). Delay-period activity in the primate prefrontal cortex encoding multiple spatial positions and their order of presentation. *Behavioral Brain Research*, 84, 203-223.

- [4] Clower, W. T., & Alexander, G. E. (1998). Movement sequence-related activity reflecting numerical order of components in supplementary and presupplementary motor areas. *Journal of Neurophysiology*, 80, 1562–1566.
- [5] Dominey PF, Arbib MA, Joseph JP (1995) A Model of Cortico-Striatal Plasticity for Learning Oculomotor Associations and Sequences. *J Cog Neuroscience*, 7:3, 311-336
- [6] Dominey PF (1995) Complex Sensory-Motor Sequence Learning Based on Recurrent State-Representation and Reinforcement Learning. *Biological Cybernetics*. 73, 265-274
- [7] Maas W, Natschläger T, Markram H (2002) Real-Time Computing Without Stable States: A New Framework for Neuroal Computation Based on Perturbations. *Neural Computation* 14, 2531-2560
- [8] Jaeger, H. (2003). Adaptive nonlinear system identification with echo state networks. In *Advances in neural information processing systems* (pp. 593–600).
- [9] Verstraeten D., B. Schrauwen, M. D’Haene, D. Stroobandt, An experimental unification of reservoir computing methods, *Neural Networks* 20 (3) (2007) 391–403.
- [10] Dominey PF (1998a) Influences of Temporal Organization on Transfer in Sequence Learning: Comments on Stadler (1995) and Curran and Keele (1993), *J Exp Psychology: Learning, Mem and Cog*, 24, 1, 234-248
- [11] Dominey PF (1998b) A shared system for learning serial and temporal structure of sensorimotor sequences? Evidence from simulation and human experiments. *Cog Br Res*, 6, 163-172.
- [12] Dominey P.F., Ramus F., (2000), Neural Network Processing of Natural Language: I. Sensitivity to serial, temporal and abstract structure of language in the infant. *Lang. and Cognitive Processes*, 15(1) 87-127
- [13] Dominey PF, Lelekov T, Ventre-Dominey J, Jeannerod M. (1998) Dissociable processes for learning the surface structure and abstract structure of sensorimotor sequences. *J Cogn Neurosci*. 10(6):734-51
- [14] Douglas, R. J. & Martin, K. A. Neuronal circuits of the neocortex. *Annu. Rev. Neurosci.* 27, 419–451 (2004).
- [15] Dehay, C. & Kennedy, H. Cell-cycle control and cortical development. *Nature Rev. Neurosci.* 8, 438–450 (2007).
- [16] Grossberg S, Pearson LR. (2008) Laminar cortical dynamics of cognitive and motor working memory, sequence learning and performance: toward a unified theory of how the cerebral cortex works. *Psychol Rev.* Jul;115(3):677-732.
- [17] Tanji J, Shima K, Mushiake H (2007) Concept-based behavioral planning and the lateral prefrontal cortex. *Trends Cogn Sci.* Dec;11(12):528-34
- [18] McDonnell MD, Abbott D (2009) What Is Stochastic Resonance? Definitions, Misconceptions, Debates, and Its Relevance to Biology. *PLoS Comput Biol* 5(5): e1000348. doi:10.1371/journal.pcbi.1000348