1993-01

# Learning Temporal Contexts and Priming-Preparation Modes for Pattern Recognition

Banquet, Jean-Paul

Boston University Center for Adaptive Systems and Department of Cognitive and Neural Systems

# LEARNING TEMPORAL CONTEXTS AND PRIMING-PREPARATION MODES FOR PATTERN RECOGNITION

Jean-Paul Banquet and Jose L. Contreras-Vidal

**January 1993**

**Technical Report CAS/CNS–93–007**



Boston University Center for Adaptive Systems and
Department of Cognitive and Neural Systems
111 Cummington Street
Boston, MA 02215

# Learning temporal contexts and priming-preparation modes for pattern recognition

**Jean P. Banquet**[1] **and José L. Contreras-Vidal**[2]
Center for Adaptive Systems and Cognitive and Neural Systems Department
Boston University
111 Cummington Street, Boston, MA 02215

**Abstract**

The system presented here is based on neurophysiological and electrophysiological data. It computes three types of increasingly integrated temporal and probability contexts, in a bottom-up mode. To each of these contexts corresponds an increasingly specific top-down priming effect on lower processing stages, mostly pattern recognition and discrimination. Contextual learning of time intervals, events' temporal order or sequential dependencies and events' prior probability results from the delivery of large stimuli sequences. This learning gives rise to emergent properties which closely match the experimental data.

# 1 INTRODUCTION

Basic questions of information and cognitive processing theory concern the where, when and how's of spatio-temporal context processing and the impact of this contextual information on pattern recognition and event identification. Natural or artificial pattern recognition either temporal as in language or music, or spatial as in vision, character-recognition, but also motor sequence planning and robotics closely depend on contextual expertise.

In cognitive psychophysiology and neurophysiology, spatio-temporal contexts are essential aspects of complex stimuli in echological situations. They allow the desambiguation of single events. More surprisingly, they are also essential factors of simple associative learning, as in classical and operant conditioning.

In this work the term context is used in its more general acception of any type of relationship between two events or stimuli, be it spatial or temporal, deterministic, probabilistic, or syntactic. For the sake of simplicity and manipulability a probability context (probability relations between events) was investigated; but we claim that the results generalize to any type of context. A probability context in itself is neither spatial nor temporal. But because the stimuli were delivered as Bernoulli sequences we were confronted by two essential aspects of brain temporal processing: timing and temporal order. These questions have been adressed either experimentally or theoretically [1], [2], [8], [9]. This work is an attempt to bring together the most integrated results of psychophysiology and neurophysiology to build a complex, parsimonious, still biologically plausible neural network system.

In the vast repertoire of available technologies used in the neurosciences, electrical or electromagnetic brain imaging, particularly event-related potentials (ERPs), present several advantages: (a) An unmatched time-resolution power allows a monitoring of cognitive processes in quasi real-time, at the functional level; (b) a large scalp coverage by electrodes allows the location of the brain generators of the information processing activities [1] [2]. As such ERP experiments provide information about how, when and where neuronal assemblies cooperate during cognitive tasks. Thus a level of brain organisation is probed intermediate to that of neurobiology and behavior. This level is congruent to that of artificial neural networks, which also probe neural populations, still preserving relevance to the functional level, thanks to emergent macroproperties.

A dual task involved (a) an explicit discrimination between high and low pitch auditory stimuli presented in Bernoulli series, necessary and sufficient to elicit a go-nogo motor response; and (b) an implicit automatic time interval, temporal order and probability contexts evaluation embedded in an incidental learning paradigm, assessing both medium-term and long-term learning. The experiment allowed to asses "off-context" and "in-context" cognitive processing . Because of the sequential stimulus delivery, and the fixed ISI (interstimulus interval), this paradigm is also a formal analog of a temporal conditioning.

[1]Supported by INSERM, and DGA/DRET Grant # 911470/A000/DRET/DS/DR
[2]On leave from Monterrey Institute of Technology, México, and by CONACYT # 63462.

The experimental results [1], [2], [3] will be reported only as far as they back-up the architecture of the model . They determine the main features of the system and strongly constrain the choice of the parameters.

# 2 EXPERIMENTAL RESULTS and SYSTEM DESIGN

From the experimental results, three types of temporal contexts can be individualized, corresponding to increasing levels of integration, but not necessarily of complexity, in the structure of the input data. To each of these contexts correspond a particular type of top-down priming, which is more or less specific according to the context that it comes from.

## 2.1 Timing Module

Constant temporal relations between stimuli induce time evaluation of the interval between them. This type of contextual information gives rise to a nonspecific preparation affecting uniformly the different categories of stimuli. The timing module (Figure 1a) is based on the same principle as in Grossberg and Merril [10]; a battery of 30 nodes, each reacting to a stimulus with a specific dynamic, so as to cover the entire spectrum of biologically plausible intervals. Each stimulus has a double valence:

1. It sends a Now Print (NP) signal to the LTM (Long Term Memory) weights, that induces the learning and memorisation of the spectral activation of the nodes in the battery. The nodes maximally active at the time of the NP induce maximal learning at the corresponding synapses. In this way, NP acts as a time marker for the end of the previous interval which is evaluated and learned.
2. The same stimulus sends a Start signal that reactivates the battery of nodes. As such it acts as a time marker for the beginning of the next interval.

## 2.2 Temporal Order Module

Sequential information is recorded in short-term (STM) and subsequences are categorized as chunks. Prototypical subsequences (so called because the local probability of their events agrees with the prior probability of the entire sequence) are predominantly learned because they constitute about two thirds of the total stimulus delivery time.

1. In STM, items and their repetition are represented by individual nodes, as in Bradski, Carpenter and Grossberg [5]. The first layer of the module, a repeat layer sorts out repeat items into separate channels and usher them to a winner-take-all competitive field.
2. The next layer is the temporal order proper. There item's order is encoded as a gradient of activation across the nodes [5]. Two parametric variations allow for the implementation of either a primacy or a recency gradient. The output of this layer goes both to the chunk layer and the probability module.
3. The chunk layer (similar to a F2 layer of an ART model, but devoted to categorizing chunks instead of single events) has as many nodes as possible arrangements of the stimuli of a STM span (6-7 stimuli). As such, STM limited capacity avoids a combinatorial explosion in the number of possible chunks. Subactivation of these chunks by ongoing sequences gives rise to a specific and more adapted top-down priming to the categorization module than will do activation of a probability node.

## 2.3 Probability Module

Probability context effects manifest both very early ($\leq$ 10 stimuli) at the level of the P300 component of the ERPs , and very late (after 200 stimuli) at the level of the N200 ERP component. For this reason it is assumed that the P300 system is the real processor of the probability context, while the N200 system activity is only modulated by this contextual information.

The probability network is a leaky integrator receiving inputs from the temporal order module. Its nodes are characterized by dynamics one order of magnitude slower than that of the categorisation layer of ART.

The "instar" connections from the temporal order module are adaptative weights with even slower dynamics (two orders of magnitude).

This output is fedback to the categorisation module and to the temporal order module. It gives rise to a specific priming in favor of the most frequent stimulus.

### 2.4 Categorization Module

This module is both at the origin and at the endpoint of the processing chain: – at the origin, because by its function of input categorizer it feeds information to the different contextual modules; - at the endpoint, because it receives modulatory feedback inputs from all three of them. It is based on the classical ART2 or ART3 [6] categorizers, selected because of their acceptance of analog inputs, and their depletable neurotransmitter synapses.

## 3 SIMULATION RESULTS

**Timing module**: The output of this module results from the summation of all the products of the spectral node activities by the corresponding synaptic weights. The timing module learns the ISI after a few stimuli. Then, the output of the module is a skewed curve, with an anticipatory response, and a peak at the expected time of stimulus delivery, even in case of stimulus omission. Different time intervals (here 250 to 2000 msec) can be evaluated (Figure 2). The output of the timing module sends a nonspecific activation-priming signal to the nodes of the categorisation net.
**Temporal order and probability context**: The results presented here come from the simulation of the bottom-up mode of functioning of the system shown in Figure 1b.

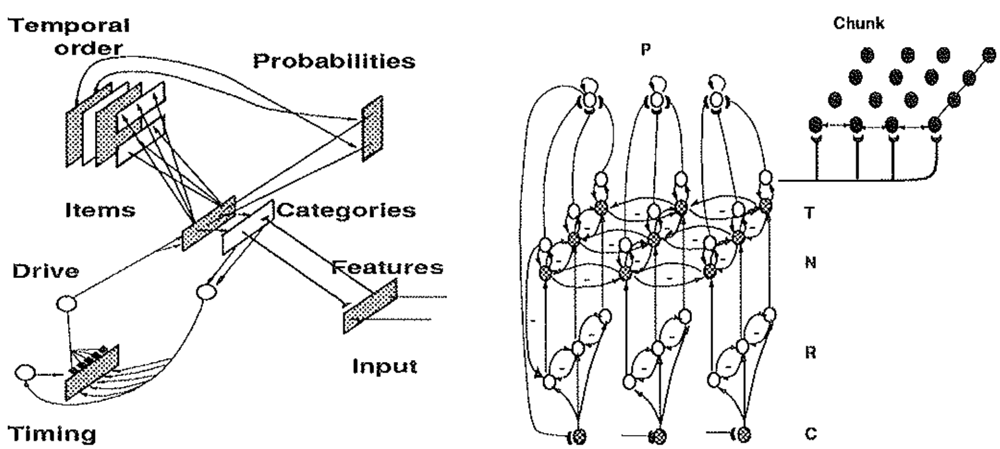


Figure 1: (a) Integrated system of five neural modules for categorization, abstraction, temporal order, probability context processing, and timing. (b) A structural neural network representation of temporal order and probability modules of Fig. 1a, with the most important intra-module and inter-module connections represented for the different layers: C Category, R Repeat, N Normalisation, T Temporal Order, P Probability.

Figures 3 and 4 show a simulation of the model described above excluding the timing module. The simulation consists of computing temporal order and context probability for the type of sequences depicted in Figure 3 **Inputs**.

Figure 3 (**Repeat**) shows the response of the repeat layer to 500 stimuli. Panel **A** and **B** depict the activity of the repeat nodes coding patterns for category A; and **B** respectively. The likelihood for a given node in the repeat layer of being selected depends on the current state of activation of the network and on the input sequence. Each black line represents an input to the network. Dotted lines represents the short-term reverberation of repeat activities. A different node is selected at each new input during the time span of a traveling window; The inhibition of the repeat nodes by the temporal order layer is suppressed

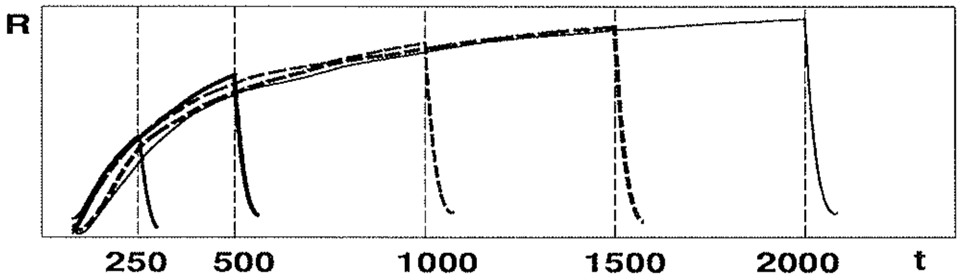


Figure 2: Output of the timing module as a function of ISI's ranging from 250 to 2000 msec. After learning these different ISI's, the output of the module anticipates the expected occurrence of the input, and presents a maximal activity at the time of this occurrence even in the case of input omission.

by the removal of the corresponding items off the TOM. Hence previously used repeat channels can become again available. In addition, there is only one highly active node in this layer at any moment due to the competitive dynamics among nodes of the same category.

The **TOM** panels of Figure 3 depicts the activity of the temporal order network for category **A** and **B**. These layers register the order of the input patterns by amplitude coding in a 2D array (items *times* repeats). The input category is encoded by spatial coding in this case, of two categories, but the system can encode the order of an arbitrary number of categories. The temporal order show an effect of recency, such that recent items have higher activity in this layer than previous items.

The activation level of the probability layer (Figure 4) reflects the local frequency of events. Note that the activation level for each of the two categories track the local probability. For unequal probabilities (Fig. 4a), two types of dynamics are superimposed in the output of this module:

(a) Temporal order or sequential dependency effects manifest as fluctuations of the graph; (b) local probability effects show up after a few stimuli coded by the level of ampkitude of the 2 curves; and (c) the asymtote level of the graph reached after about 200 stimuli ($t(r)$) corresponds to a modulation of the node activity by LTM synaptic weights, and reflect prior probability effects.

Time $t(s)$ represents a switch to an equal probability of events. Note that the probability equilibrates at a level of 0.5 representing the true probability of the last 100 trials. Note also that the event with lower probability produces more local variations in its probability node. Figure 4b show the probability curves in the case of equal input probabilities.

**References**

[1] Banquet, J.P. and Grossberg, S. (1987). *Applied Optics*, **26**, 4931-4946.

[2] Banquet, J.P., Smith, J.M., Guenther, W. (1992). *In Motivation, Emotion, and Goal Direction in Neural Networks*, Levine and Leven (Eds.) Laurence Erlbaum: New York, 169-208.

[3] Banquet, J.P., and Contreras-Vidal, J.L. (1992). In: *Proc. of the ICANN'92 International Conference on Artificial Neural Networks.* Brighton, UK, 4-7 september 1992.

[4] Banquet, J.P., and Contreras-Vidal, J.L. (1992). *In Proc. of the IJCNN International Joint Conference on Neural Networks*, Baltimore, Maryland, June 7-11, 1992. Vol.I, pp. 541-546.

[5] Bradski, G. Carpenter, G., Grossberg, S. (1991). *Proc. IJCNN.* Seattle, WA, **1**,723-78.

[6] Carpenter, G.A. and Grossberg, S.,(1990). *Neural Networks*, **3**, 129-152.

[8] Grossberg, S. (1982). In *Boston Studies in the Philosophy of Science.* **70**, 662.

[9] Grossberg, S.,(1978). In *Progress in theoret.Biol.*, **5**, Rosen and Snell (Eds.), New York: Academic Press, pp. 233-374.

[10] Grossberg, S. and Merrill, J.,(1992). *Cognitive Brain Research*,**1**, 3-38. *NeuralNetworks*, **2**, 79-102.

[11] Acknowledgements: Thanks to Gail Carpenter and Steve Grossberg for their helpful comments at different stages of this research.

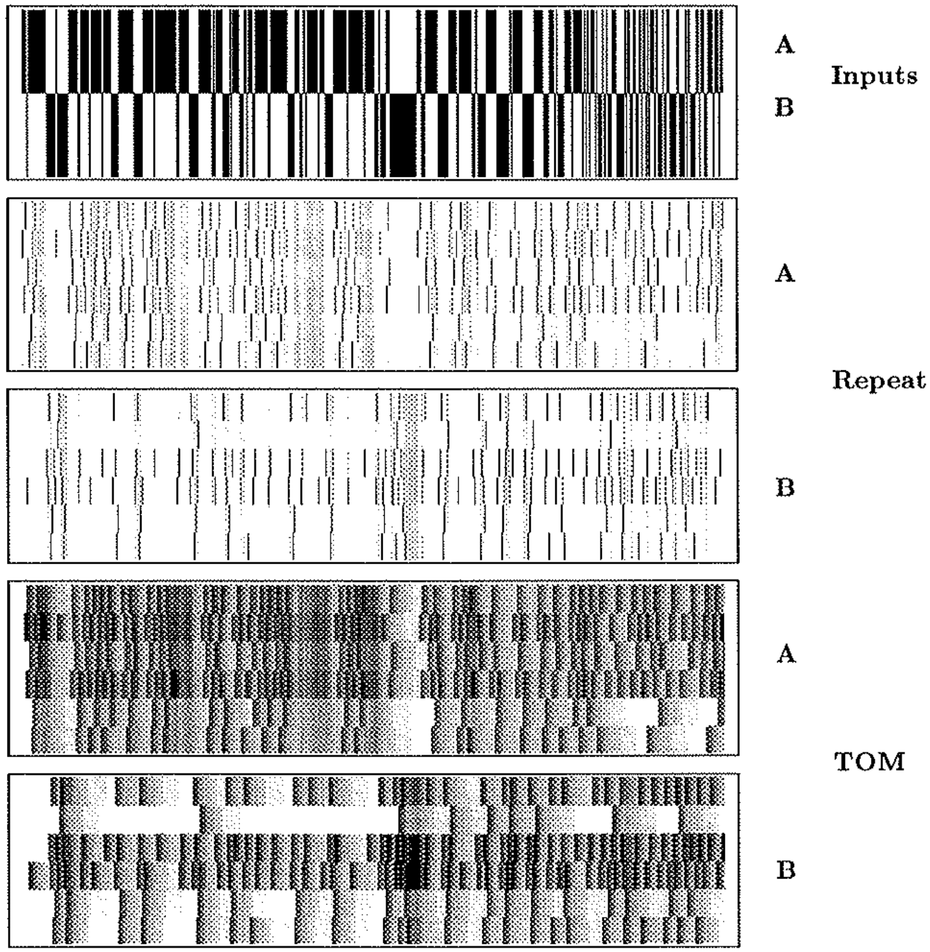


Figure 3: (**Inputs**) Sequence of 500 stimuli from categories A and B. The input sequence has a probability of $P(A) = 0.8$ and $P(B) = 0.2$ for the first 400 stimuli, then it switched to an equal probability of 0.5 for both categories for the last 100 stimuli. (**Repeat**) Repeat layer sorts out the pattern of activation according to both the probability and repetitions of the input signals for category A and B respectively and repeat items. This layer can store up to 6 repetitive items per category. (**TOM**) Activation of the temporal order network for categories A and B respectively. The figure shows a traveling window due to a combination of recency, limited short-term memory, and the temporal structure of the input sequence.

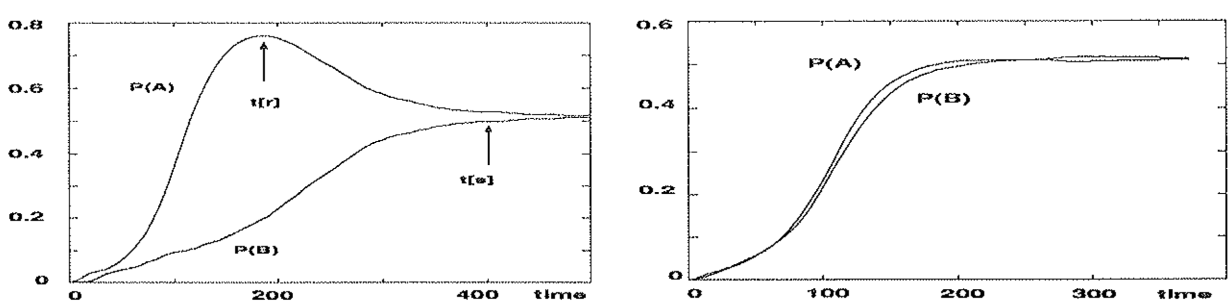


Figure 4: Activation of the probability layer. $P(A)$ and $P(B)$ represent the probability of the categories A and B respectively. (a) $P(A) = 0.8$ and $P(B) = 0.2$ during first 400 stimuli up to time $t(s)$; $P(A) = P(B) = 0.5$ thereafter. $t(r)$ is the time where a probability habituation occurs for the node coding $P(A)$. (b) $P(A) = P(B) = 0.5$ during 500 stimuli.