# Short-term memory

Emmanuel Guigon, Yves Burnod

INSERM CREARE, Université Pierre et Marie Curie
9, quai Saint-Bernard, 75005 Paris, France

Guigon E, Burnod Y. Short-term memory both as working memory and as basis for long-term memory. In: The Handbook of Brain Theory and Neural Networks (Arbib MA, ed), 1995, Cambridge, The MIT Press, pp 867-871.

# 1 Introduction

It is now generally agreed that two temporally distinct neural processes contribute to the acquisition and expression of brain functions. Transient variations of membrane potential (neuronal activity), with a time scale of milliseconds, reflect the flow of information from neuron to neuron and define the function of neuronal networks. These variations can result in long-lasting (and maybe permanent) alterations in neuronal operations, for instance through activity-dependent changes in synaptic transmission. There is now strong evidence for a complementary process, acting over an intermediate time scale (short-term memory, STM). This process is involved in performing tasks requiring temporary storage and manipulation of information to guide appropriate actions (Goldman-Rakic 1987; Baddeley 1992). Two main issues should be addressed when studying STM: (1) How is neural information selected for storage and temporarily stored in STM for future use in a temporal sequence of sensorimotor events, and how is a large amount of information buffered when its future use is not known? (2) How can a long-term memory (LTM) representation of temporal sequences of events be constructed, and how can information selected by STM process be transferred to LTM?

Fig. 1 outlines a general scheme which allows models including short-term and long-term memory to be compared. A network of units processes spatial and temporal information, represented by their short-term (milliseconds to a few seconds) activities ($x_i$,$y_j$). The units store information by changing their synaptic weights ($W$), with long-lasting effects (days, years). Information can be stored over intermediate time scales of seconds, minutes or hours by short-term memory elements, represented by units ($z_k$). At the neural level (Fig. 1A), STM appears to be an intermediate step between neuron activity and LTM within single neurons or simple circuits. At the system level (Fig. 1B), two brain structures have been extensively studied for their role in STM processes, the hippocampus (Squire et al. 1993) and the prefrontal cortex (Fuster 1989).

The present article discusses the relationships between neuronal activity, STM and LTM at the neural and system levels, and then presents a neural network model, which illustrates the properties of short-term memory in the prefrontal cortex.

# 2 Short-Term Memory at the Neural Level

This section focuses on the temporal patterns generated by single neurons, simple circuits or networks, which may be responsible for short-term memory.

## 2.1 Biological Neurons and Simple Circuits

From Activity to STM: A wide variety of temporal patterns of activity are actively generated by neurons and local circuits of neurons, such as transforming transient inputs into long-lasting sustained or oscillatory activity. Experimental studies in invertebrates (Harris-Warrick and Marder 1991) have demonstrated that such temporal patterns produce motor programs and are generated both by the molecular properties of each neuron, and the connectivity of the local network.

It is now well established that, in vertebrates, long-lasting activities are neural correlates of transient memory processes (mainly in the frontal lobe of the cerebral cortex). This pattern of activity allows past events to be represented and behavioral reaction to future, predictable events to be prepared (Goldman-Rakic 1987; Fuster 1989).

Several models have addressed the question of how to maintain such a sustained activity. One popular model is a network with reverberating excitatory feedback loops (Hebb 1949). A model of reciprocally inhibitory neurons, described by a system of nonlinear differential equations, can generate bistable activities (Kirillov et al. 1993). Such a model describes the conditions of stability of the two states, the role of noise and the input commands for transitions between states. Zipser et al. (1993) provided direct evidence for bistability of cortical neurons in a recurrent neural network trained to mimic the input-output characteristics of an active memory module.

The intrinsic properties of a single neuron can also be responsible for generating bistable activities, via a set of ionic channels (for example, persistent sodium currents in the spinal cord): one state is silent, and the other is continuous activity that can be triggered by a transient synaptic input (Harris-Warrick and Marder 1991).

**From STM to LTM** A cascade of molecular events occurs in neurons after synaptic activation; these include the activation/inactivation of the various types of sodium and potassium channels with different time constants, a calcium influx and second messenger cascades, short-term changes in the probability of transmitter release, and short-term potentiation (STP). These events define memory traces which outlast the duration of synaptic events. They also constitute initial steps for the formation of longer memory traces, such as long-term potentiation (LTP) which can last for hours (T.H. Brown, cross-reference).

## 2.2 STM in Neural Network Models

From a computational point of view, a simple way to implement short-term memory is to consider that neuronal variables have an effect which outlasts their duration. This effect can concern synaptic weights, synaptic inputs, or membrane potential as illustrated in Fig. 1A.

Short-term memory can appear to be an intermediate step in the learning process at the level of each synapse, in the same way as STP and LTP (Fig. 1A1). The strength of the synapse is transiently modulated by the successive events in a sequence. Since associative learning rules, such as the wide classes of Hebbian rules, are based on the temporal coincidence of two events, transient synaptic modifications allow the time overlap to be increased, and thus association between temporally separate events to be learned. Sutton and Barto (1981), proposed a model in which the time scale of neuronal activations (milliseconds to hundred of milliseconds) is extended to the time scale of temporal correlations between successive sensory and motor events during classical conditioning (seconds to minutes).

Temporal sequences of events can be turned into spatial patterns ((Fig. 1A2). In Time Delay Neural Networks (Waibel 1989), a sequence is represented by a vector in such a way that the position in the vector corresponds to the temporal order of events. Thus, the events occurring in a pre-selected time-window can be learned as a single spatial pattern.

Time can also be processed when units are linked by recurrent connections (Fig. 1A3). In recurrent networks, a subset of input neurons represents the trace of activity of output neurons that is the result of the computation performed by the network in the previous time step (J.L. Elman, cross-reference). A new input is thus interpreted in a specific context, which can be learned from the previous computations performed by the network. In this way, the neuronal operations at a given time are modulated by the recent history of the network. In these models, memory results from sequential transitions between states rather than from the buffering of specific items.

# 3 Short-Term Memory at the System Level

Short-term memory may also be a property of large-scale neural architecture involving cortical and subcortical regions of the brain (Fig. 1B).

## 3.1 Hippocampus

The hippocampus is important for STM to LTM transfer, but it does not appear to be the substrate of LTM (Squire et al. 1993). Monkeys with hippocampal lesions are severely impaired at remembering recently learned objects, but perform correctly with objects learned long ago. The hippocampus is needed for a short period of time after learning. Permanent memory develops in adjacent cortical areas of the temporal lobe, to which the hippocampus is connected in parallel (see Fig. 1B1). The neural correlates of the long-term memory storage of new patterns in these temporal regions are now well documented: for example, in recognition of visual patterns, neural activities are selective for “prototypes”, invariant in size and orientation, and sustained activities represent the temporal links between these prototypes.

Memory processes in the hippocampus appear to be based on three different forms of plasticity within a serially organized anatomical circuit that comprises the cortico-hippocampal pathways (entorhinal cortex $\rightarrow$ dentate $\rightarrow$ CA3 $\rightarrow$ CA1). Experimental data indicate that synaptic potentiation can persist for hours in mossy fibers (between dentate and CA3), for several days in cortical projections to the dentate gyrus, for several weeks in the CA1.

The hippocampus can be viewed as a control system between STM and LTM in temporal cortical areas (Carpenter and Grossberg, cross-reference). This “search and orienting” system is able to determine whether an input is a new example of a previously stored prototype or a new prototype. A “resonant state” appears when low-level inputs and high-level expectancies are matched. During this state, the input example can be stored. When there is a mismatch, the hippocampal control system triggers a memory search for a better category by activating a new high level expectancy (a new “hypothesis”). If the input is too different from any previously learned prototype, the hippocampal control system selects an uncommitted population of high level neurons to store a new category. Once a memory is formed, the hippocampus is not needed for retention or retrieval: familiar events have a direct access to their recognition code.

## 3.2 Prefrontal Cortex

The prefrontal cortex is involved in integration of temporally separate events into purposive behavioral structures (Goldman-Rakic 1987; Fuster 1989). This function is reflected in the performance of tasks using temporal delays for structuring behavioral reactions to environmental stimuli. The paradigmatic test is the delayed response task, which requires a subject to memorize an instruction stimulus and to wait for a go signal before responding to it. This task is typically impaired after lesions of the prefrontal cortex (Fuster 1989).

Prefrontal neurons recorded during delayed response tasks in monkeys display patterns of sustained activity which reflects the short-term mnemonic aspects related to instruction cues, the expectation of forthcoming signals and the preparation of the behavioral reaction. Sustained activities have three important characteristics, which define their cardinal role in the learning and execution of behavioral tasks. First, whatever the modalities used (visual or auditory cues, arm or eye movement responses), they occur during the delay between an instruction cue and the final permission to use the information contained therein to produce a response. Second, the duration of the activity is linked to the duration of the delay. Increasing the delay's length leads to a prolonging of the activity. Third, these activities are a product of learning and there appears to be a relationship between the amount of delay activation and the level of performance (Fuster 1989).

Goldman-Rakic (1987) has proposed that the prefrontal cortex is necessary for expression of behaviors guided by representation or internalized models and not when the behavior is guided by external stimuli. The mechanism of the prefrontal cortex is related to a distributed system of interconnected neural networks. Specific functions would thus come from dynamics of the system and interactions between independent networks, rather than from a strictly hierarchical processing based on the convergence through association regions (Fig. 1B2). In a such a system, the "working memory", defined as the formation of selective memory traces of relevant events, appears as a relevant concept which characterizes the specificity of prefrontal functions (Goldman-Rakic 1987). Short-term memory in the prefrontal cortex appears to be subserved by sustained activities. Furthermore, these activities may also be involved in the formation of permanent memory (Fuster 1989). Thus the same mechanism is likely to participate to formation and retention of memories. This property is in contrast with the hippocampus, in which different mechanisms contribute to the formation and the retention of memories.

## 4 A Model of Short-Term Working Memory

We have built a computational model of prefrontal circuits to illustrate a strategy for implementing short-term memory in a neural network (Guigon et al. 1995). Based on the principles of organization and operations in the prefrontal cortex, this model shows that short-term memory in a neural network can be obtained by processing units which switch between two stable states of activity (bistable behavior) in response to synaptic inputs. The sustained activity of a given neuron represents a temporal link between two sensory or motor events. It also shows that long-term representation of tasks requiring short-term memory can result from activity-dependent changes in the synaptic transmission controlling the bistable behavior. After learning, the sustained activity of a given neuron represents both the selective memorization of a past event and the selective anticipation of a future event.

## 4.1 Description of the Model

The neural network model, designed according to the principles of organization of prefrontal connections, was trained to execute a delayed response task (Fig. 2). The architecture of the network is shown in Fig. 2A. Each sensory event is coded by the all-or-none activation of a specific unit in the sensory layer, and movements towards the levers are coded in the motor layer. Matching units model neurons in the associative sensory and motor areas connected to the prefrontal cortex. These units implement sensorimotor relations, such as a direct relation between the position of the lever and movement toward the lever. Bistable units model prefrontal neurons. Each bistable unit is reciprocally connected to one matching unit and receives non-reciprocal projections from some other matching units. This connectivity defines multiple interactions between matching and bistable units, but does not correspond to an a priori representation of particular functions. Bistable units are connected to a drive pathway $d$ (thirst), which is made active at the beginning of each behavior of the network, and to a reinforcement pathway $r$ (receipt of liquid), which is activated when a correct behavior is produced by the network.

The function of the network is defined by the dynamics of processing units and by the adjustable connection coefficients between processing units. Neural processing function of matching units is modeled by a nonlinear interaction between inputs, which reflects the modulation of sensory inputs and motor outputs by memorized conditions. Matching units respond to the coactivation of sensory and bistable inputs, but not to activation of either input alone. We postulate that prefrontal neurons have two stable states of activity (bistable), and that transitions between these states are elicited by synaptic inputs (Fig. 2B; Guigon et al. 1995. We also postulate that this bistable behavior is controlled by learning and allows sensorimotor sequences to be built up under the control of a reinforcement signal.

## 4.2 Bistable Units Implement Short-term Memory

Computer simulations of the neural network in Fig. 2 were used to train it to execute a delayed response task in three successive stages (1: movement, reward; 2: go signal, movement, reward; 3: instruction stimulus, go signal, movement, reward). The rationale for this protocol is that the training protocols used with animals are progressive, stage by stage procedures. The contribution of bistable units to the execution of the delayed response task is illustrated in Fig. 3A. Each graph displays qualitatively the activity of three bistable units at a given training stage. During execution of the task (stage 3), bistable units display different patterns of activity defined by the temporal relationship between task events and peaks of activity (Fig. 3A). Each unit is active between two successive task events. The most interesting pattern is the differential delay activity. This is a sustained activity between the onset of the instruction stimulus and the onset of the go signal specific for right vs left trials. All these patterns have been described in the prefrontal cortex during the delayed response task (Fuster 1989).

At each training stage, bistable units play a complementary role in encoding the temporal structure of the task. Individual units are selective for a specific sequence of events, but the set of units is able to bridge all the gaps between the events of the current task. Matching units displayed transient activity that was time-locked to sensory or motor events and that was correlated with the end of activity in bistable units. They signal the occurrence of a specific sensory or motor events in the context of a specific behavior.

### 4.3 Long-term Changes in Bistable Units

Variations in the activity of bistable units are correlated with the changes in reinforcement contingency, depending on variations in the reinforcement rate (Fig. 3B). Two behaviors are alternatively performed by the network when changing from stage 1 to stage 2: one is the previously correct behavior (self-initiated movements) and the other is the new correct behavior (stimulus-triggered movements). The mean activity during reinforced trials increases for leftward self-initiated movements during the first stage. During the transition from stage 1 to stage 2, activity first decreases and then increases with the increase in the performance rate. The same phenomenon occurs between stage 2 and stage 3 (Fig. 3B). The experience gained at each trial in the learning period is thus transferred to a long-term representation of the task.

## 5 Discussion

In most neural networks, information stored into long-term memory reflects correlation between transient neuronal activities. This form of storage is efficient for encoding long-term memory of objects, but less efficient for encoding temporal sequences of events (Sutton and Barto 1981). We have described different strategies, which use short-term memory mechanisms to link temporally separate events. In some cases, memory is an implicit consequence of neural network architecture and neuronal dynamics. In other, an explicit neural correlate of short-term memory is observed (theta rhythm, sustained activity). We have presented a simple model in order to illustrate a possible mechanism of short-term working memory in the prefrontal cortex. The model has shown the dual role of sustained activity in the short-term retention of relevant cues and in the formation of long-term memory of simple sequential behavior. Further study of these strategies should provide initial cues for the understanding of complex behaviors involving planning, reasoning, language.

## References

Baddeley A (1992) Working memory. Science 255:556–559.

Fuster J (1989) *The Prefrontal Cortex. Anatomy, Physiology, and Neuropsychology of the Frontal Lobe*. New York: Raven.

Goldman-Rakic P (1987) Circuitry of primate prefrontal cortex and regulation of behavior by representational memory. In: *Handbook of Physiology, Sect 1: The Nervous System, Vol V: Higher Functions of the Brain, Chp 9* (Plum F, ed), pp 373–417. Bethesda, MD: American Physiological Society.

Guigon E, Dorizzi B, Burnod Y, Schultz W (1995) Neural correlates of learning in the prefrontal cortex of the monkey: A predictive model. Cereb Cortex 5:135–147.

Harris-Warrick R, Marder E (1991) Modulation of neural networks for behavior. Annu Rev Neurosci 14:39–57.

Hebb D (1949) *The Organization of Behaviour*. New York: Wiley.

Kirillov A, Myre C, Woodward D (1993) Bistability, switches and working memory in a two-neuron inhibitory-feedback model. Biol Cybern 68:441–449.

Squire L, Knowlton B, Musen G (1993) The structure and organization of memory. Annu Rev Psychol 44:453–495.

Sutton R, Barto A (1981) Toward a modern theory of adaptive networks: Expectation and prediction. Psychol Rev 88:135–170.

Waibel A (1989) Modular construction of time-delay neural networks for speech recognition. Neural Comput 1:39–46.

Zipser D, Kehoe B, Littlewort G, Fuster J (1993) A spiking network model of short-term active memory. J Neurosci 13:3406–3420.

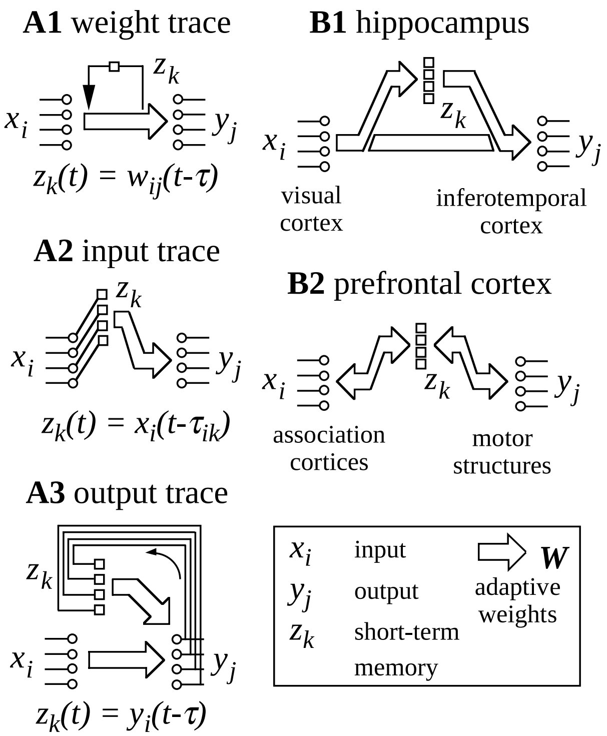

Figure 1: Architecture for short-term and long-term memory at two levels of brain organization. Each network has an input pathway ($x_i$), an output pathway ($y_j$), a short-term memory pathway ($z_k$), and a set of adaptive weights ($W$). **A**. The neural level. **A1**. STM units store transient variations in synaptic weights. **A2**. STM units store synaptic inputs in delay lines. **A3**. STM units store recent history of the activity in the network provided by recurrent connections. **B**. The system level. **B1**. STM units in the hippocampus contribute to the transfer of information from STM to LTM. **B2**. STM units in the prefrontal cortex are the temporal link between sensory and motor events.

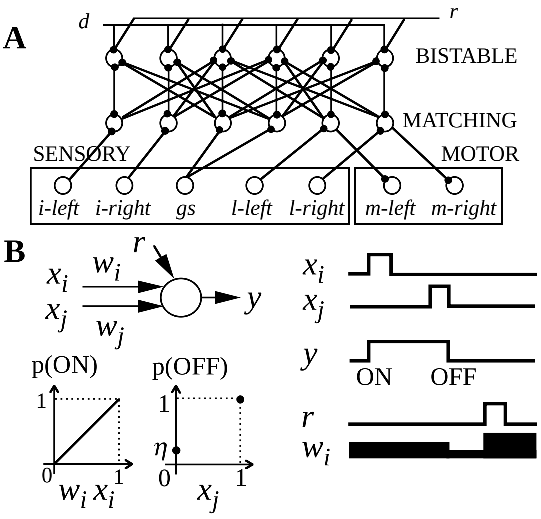

Figure 2: **A**. Architecture of the network for learning a delayed-response task. The task involves two lights mounted above two horizontally arranged levers and a trigger light. At each trial, one light (instruction stimulus) comes on for a short period; a few seconds later, the trigger light (go signal) comes on and the animal touches the lever indicated by the instruction: it receives a reward. Notations: *l-left* and *l-right*: positions of left and right levers; *m-left* and *m-right*: movements toward the levers; *gs*: go signal; *i-left* and *i-right*: instruction stimuli; *d* and *r* for drive and reinforcement, respectively. Black dots indicate synapses. *B*. Dynamics of bistable units. The unit has two weighted input pathways $x_i$ ($w_i$) and $x_j$ ($w_j$), a reinforcement pathway *r*, and an output pathway *y*. Variables are binary. Weights vary in [0,1]. Qualitative variations in the activity *y* and the synaptic weight $w_i$ when input and reinforcement pathways are activated as shown in the tracings. Unlike classical neural automata, which display transient responses to transient inputs, the present neuron remains activated (state ON) for some time after the input $i$. The neuron then returns to rest after the second input $j$ (state OFF). Transition to the ON state follows a classical law used to model the stochastic behavior of neurons (graph p(ON)): the probability of transition is proportional to the summed inputs. Transition to the OFF state has two components (graph p(OFF)): a spontaneous transition with a fixed probability $\eta$ (effect of noise), and an unconditional transition following subsequent inputs. Only the transition to the ON state is controlled by a synaptic weight.

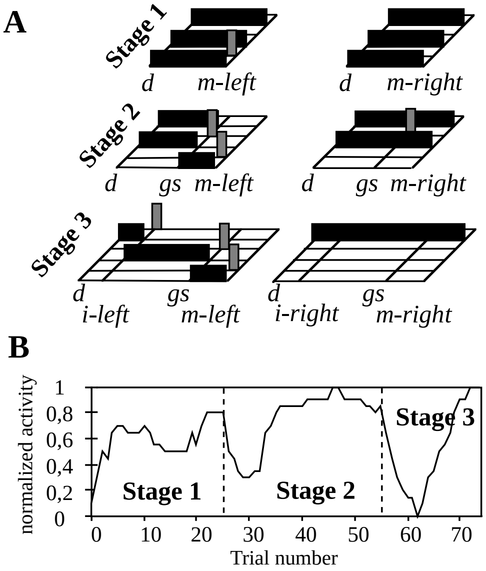

Figure 3: Computer simulations. **A**. Activities in three bistable units (*dark shaded pattern*) and three matching units (*light shaded pattern*) are qualitatively displayed for each training stage and for left and right trials. The task events are those described in Fig. 2. Note the gradual changes in the relationships between neuronal activity and task events and the differentiation for left vs right trials. **B**. Variations in the level of activity of a bistable unit during the training period. The graph is constructed from the activity during reinforced left trials. Each horizontal division corresponds to a trial. *Vertical dashed lines* indicate the transitions between training stages. Note the combination of increasing and decreasing activity: activity decreases at the transition between two stages and increases after the transition.