Dynamical effects on familiarity discrimination

J.M. Cortes, A. Greve and M.C.W. van Rossum

Institute for Adaptive and Neural Computation.
School of Informatics, University of Edinburgh
Edinburgh EH1 2QL, UK

Abstract. One process involved in recognition memory is familiarity discrimination. Familiarity distinguishes almost immediately after stimulus presentation whether the item was previously encountered (old) or novel. By using a formalism based on attractor neural networks, we discuss different dynamical processes affecting familiarity discrimination. First, we compare two different familiarity discriminators, the previously proposed energy (FamE) [1] and the temporal derivative of the energy (FamS). This second measure relies on differences in the dynamics of the network when novel or old stimuli are presented. Contrarily to FamE, FamS depends on details of the dynamics of the network. As a result, and counterintuitively, famS is enhanced by random fluctuations in the neural activity. Finally, we present some preliminary results showing how familiarity discrimination is affected by activity dependent mechanisms at synaptic level, such as short-term depression and facilitation. These results allow formulating new models regarding optimal dynamics and familiarity discrimination.

FAMILIARITY DISCRIMINATION

Recognition memory describes the ability to identify an item as previously encountered. It is mainly supported by two distinct types of retrieval, familiarity and recollection. Familiarity refers to a fast acting process that reflects a quantitative assessment of memory strength, whereas recollection is associated to the retrieval of contextual information about the event. In experimental psychology, *dual-process* theory proposes a detailed account of how recognition memory is supported by both familiarity and recollection (for review see [2]). The neural basis of both familiarity and recollection still remains unclear. However, recollection is often thought to corresponds to an attractor state in an auto-associative network [3], and recently a network for familiarity discrimination has been proposed [1, 4].

   In this paper we focus on familiarity discrimination, which has been widely debated in different classes of experiments. In the early $70's$, Standing and collaborators reported psychophysical experiments with humans [5], in which they presented a large set of arbitrary pictures for 5 seconds each. After two days most of the images were successfully recognized as previously seen. The most surprising fact was that the number of retained pictures fitted to a power law relationship with the number of presented images, which means that the capacity of recognition memory for pictures is apparently limitless. Electrophysiological studies [6, 7], using single cell recordings in monkeys

and rats, demonstrate that in perirhinal cortex about the 30 percent of neurons respond actively when a novel stimulus is presented. These neurons have been interpreted to act as novelty detectors. Finally, neuropsychological studies and neuroimaging techniques (Event Related Potentials, ERP) [8] showed that both familiarity and recollection are distinct in a number of ways, for instance familiarity is rapid (less than 0.6 seconds) whereas recollection is slow.

The high storage capacity of familiarity memory has been reproduced in computational models [1]. Under a wide range of conditions the capacity for familiarity discrimination is proportional to the number of synapses of the network. Thus the capacity of familiarity is much larger than the recollection capacity, which is proportional to the square root of the number of synapses (the number of nodes in a fully connected network) [3, 9]. Intuitively this is easily understood, familiarity needs to store just a single bit (familiar vs. non-familiar) per pattern whereas to recall an event requires retrieval of the whole pattern. In the next sections, we introduce an alternative familiarity discriminator and we show how different dynamical processes, such as time evolution, fluctuations in neural activity and short-term plasticity affect familiarity discrimination.

A NOVEL FAMILIARITY DISCRIMINATOR

We consider a network of N neurons, each one represented by its activity s_i, with $i = 1, \ldots, N$. Any two neurons connected by the synaptic weights ω_{ij}. As standard in artificial network models, the network has a learning phase in which it encodes P stimuli x_i^μ in the weights, where $\mu = 1 \ldots P$ labels the patterns. After this learning protocol the synaptic weights remain unchanged, but can be modified by short-term plasticity, which will be studied in the next section of this paper.

During the test phase, the network's performance is evaluated. At $t = 0$ the stimulus is loaded into the network $\mathbf{s} = \mathbf{x}^\mu$. The network energy, defined as $E(t) \equiv -\sum_{ij} \omega_{ij} s_i(t) s_j(t)$, at $t = 0$ is able to discriminate among old and novel stimulus. This energy is of order $-N$ for learned stimuli, whereas the energy for novel stimuli is of order 0. Consequently, the energy or familiarity for old and novel stimuli, are macroscopically different (they differ order N). This is the essence of the FamE familiarity discriminator [1]. Note that because the energy is evaluated at $t = 0$, and hence before any dynamics takes place, the discriminator is independent of dynamics, such as time evolution or fluctuations on neural activity.

However, the energy is not the only way to perform familiarity discrimination. We report here how the time derivative of FamE, namely FamS, is a familiarity discriminator as well [10]. Interestingly, this familiarity was originally proposed by Hopfield in his seminal paper [3], but to our knowledge it was not further explored . The slope, defined as $S \equiv dE/dt$, tell us how quick the network's energy changes, when either a novel or old stimulus is presented. To compute the slope we rewrite the energy as a function of overlaps among the network activity and the different stored patterns in the learning rule ($E = -\sum_{\mu=1}^{P} [m^\mu]^2$, for details see [9, 10]). The time derivative of the energy is propor-
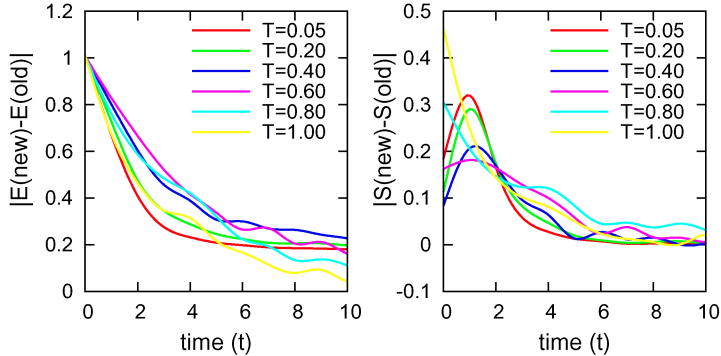
FIGURE 1. (Color Online) The time evolution of the discriminability among novel and old stimuli. Computer simulations of the time evolution of two different familiarity discriminators, energy (left) and slope (right) – see the main text for further details definitions–. For both FamE and FamS, familiarity discrimination is best briefly after stimulus presentation ($t = 0$). Contrary to FamE, the discrimination for FamS at initial time ($t = 0$) is highly dependent on the temperature parameter T. In general, the higher the temperature, the better the discrimination for FamS. For FamS, the peak occurring for low temperature nearby $t = 1$ is due to the increasing slope associated to a new stimulus. Both discriminators are represented in units of N. Parameters in the simulation are $N = 1600$ neurons and $P = 40$ number of uncorrelated patterns.

tional to the time derivative of the overlaps[1], and can be computed by using standard mean field analysis [9]. The time evolution of the discriminability, defined as the difference among responses associated with novel or old stimuli, for both FamE and FamS is depicted in Fig. 1. Immediately after stimulus presentation, both discriminators work correctly. However, after a short transient, both familiarity discriminators deteriorate. The slope tends to zero as time progresses because the network evolves towards a fixed point and becomes stationary. Also the energy difference tends to zero. Although measures to discriminate old and novel stimuli once an attractor state has been reached have been proposed [11], the performance of such measures is generally poor.

Fig. 1 illustrates further the effect of neural activity fluctuations on familiarity discrimination. These fluctuations are controlled by the so called temperature T, a parameter which controls the random transitions of the nodes [9]. After the stimulus presentation, larger fluctuations of the neural activity ($T > 0.50$ in Fig. 1) improve FamS. This fact can be intuitively understood (for a detailed study, see [10]). The energy and its time derivative can be separated into signal and noise contributions. The signal for the slope is proportional to the rate of change of the energy, and therefore proportional to

---

[1] By definition of the slope, time derivative of the energy equals $S = -\sum_{\mu=1}^{P} 2m^{\mu} dm^{\mu}/dt$. We used a mean field approach to obtain dynamical equations for the overlaps. In general, they have the form $dm^{\mu}/dt = \Phi(\mathbf{m}, T)$, being $\mathbf{m} = \{m^{\nu}\}_{\nu=1}^{P}$ the vector of overlaps. The function Φ depends on the weights, the total number of patterns stored (P) and the correlations among them.

the rate of change of the overlap between the network activity and the stimulus. At low temperatures, the signal associated to an old stimulus is very low as the overlap with the stimulus is almost invariant. Contrarily, at higher temperature, the overlap for the old stimulus changes drastically (decays from 1 to 0) and consequently the slope-signal contribution increases considerably (the higher temperature, the higher signal for the slope discriminator). The noise component for the slope, although dependent on T, is similar for both old and novel stimuli. As a result the main temperature dependence stems from the signal term. In contrast, by definition FamE at initial time is independent on neural activity fluctuations.

For FamS and low temperature, discrimination increases over time (corresponding to the peak in Fig. 1, right) due to the increasing slope associated to the new stimulus. Initially, all the overlaps among the network and the different stored patterns are 0 for both new and old stimulus. As time progresses the slope for the old stimulus does not move but stays nearby the stable fixed point, i.e. slope 0, In contrast, with a novel stimulus, some of the overlaps increase during a short transient (this is not satisfied within the high temperature regime), before it goes to zero as well as it reaches an attractor. This behavior is also illustrated in the Signal to Noise Ratio illustrated by Fig. 2. Analytical calculations in agreement with the simulations presented in Fig. 2 will be published elsewhere [10]. For very high temperature, as reported the SNR illustrated in Fig. 2, the efficiency for FamS becomes similar to FamE. The analytical expressions show that, first, for FamS the storage capacity depends on fluctuations on neural activity, and second, that there are mainly two different regions with different storage capacity. In absence of fluctuations, capacity is minimum and scales with square root the number of synapses. For very high fluctuations, storage is maximum, and again scales with the number of synapses (the same as by using FamE [1]).

DYNAMICAL SYNAPSES AND FAMILIARITY DISCRIMINATION

Synaptic plasticity shapes neuronal information processing on multiple time scales [12]. Here we include dynamical synapses in the model. Although the basal strength of these synapses is fixed after learning, the actual strength is modulated by presynaptic activity. Using a standard model of short-term depression and facilitation [13], we recently reported theoretical and computational implications of dynamical synapses including both short-term depression and facilitation on the recall processes in ANNs [14]. In a similar manner, we define the dynamical synaptic weight connecting the presynaptic neuron j with the postsynaptic i as $\omega_{ij}^{\text{dynamic}}(t) = \omega_{ij}^{\text{static}} R_j(t) U_j^{\text{eff}}(t)$, where R_j and U_j^{eff} are dynamical processes which depends on $s_j(t)$, the presynaptic neural activity. R represents a positive fraction of neurotransmitters in a recovering state from synaptic depression. U^{eff}, analogously to the release probability in the quantal model [15], describes the increase of effective use of neurotransmitters due to facilitating mechanisms. When both $R_j(t) = 1$ and $U_j^{\text{eff}}(t) = 1$ for all neurons, we retrieve the static synapse case.

We report here some preliminary results on how FamE is affected by dynamical
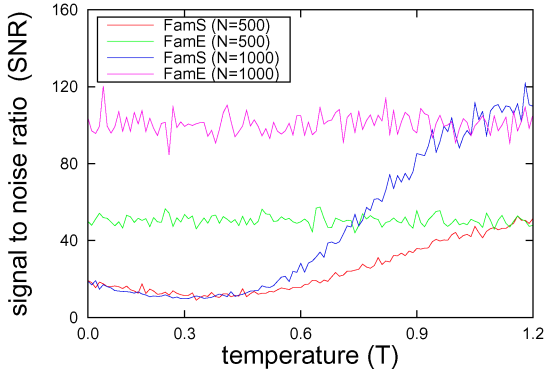
FIGURE 2. (Color Online) The influence of neural activity fluctuations (controlled by T) on both energy and slope familiarity discriminators. We compute the Signal to Noise Ratio as a function of the temperature parameter. After stimulus presentation, $t = 0$, discrimination based on energy is independent of temperature, while slope based discrimination depends strongly on temperature. For high temperature, and fixed number of patterns, discrimination based on slope performs equally the energy. We simulated two networks of $N = 1000$ and $N = 500$ neurons and $P = 50$ uncorrelated patterns. For each fixed value of temperature, the Signal to Noise Ratio has been computed by averaging over 100 different realizations of patterns.

synapses, and its relation with the storage capacity. By using Signal to Noise Ratio studies (details will be published elsewhere), the maximum number of stored patterns that can be embedded in the network within familiarity discrimination is, after dynamical synapses, $P_{\max} = \Psi(x)N^2$, with $\Psi(x) = (1+x)^2 / \left[2 \left(1+x^2 \right) \right]$. For the situation of static synapses and FamE, the maximum storage was reported as $P_{max} = N^2$ [1]. By comparison of the two results, the function $\Psi(x)$ can be interpreted as correction on the capacity after including short-term plasticity and is illustrated in Fig. 3. The variable x depends on the dynamical synapses details and is defined as $x \equiv \Delta DEP / \Delta FAC$, where Δ denotes the rate of change after the stimulus presentation in the synaptic weight and due to presynaptic spikes for both depression and facilitation. Therefore, $x > 0$, and $\Psi(x)$ takes values ranging in $(0.50, 1.00)$, c.f. Fig. 3. The maximum value is at $[x, \Psi(x)] = [1.00, 1.00]$, which corresponds to the static synapses case [1]. The case $x \gg 1$ defines a highly depressing synapse (phase D-H), whereas $x \ll 1$ a very high facilitated synapse (F-H). Counter-intuitively, these two extreme limits have the same storage, $P_{max} \approx N^2/2$. Among these two opposite situations, the synapse has both low facilitation (F-L, $x < 1$) or depression rates (D-L, $x > 1$).

The asymmetry with respect to $x = 1$ is due to the differences regarding storage among depression ($x > 1$) or facilitation ($x < 1$). The maximum storage can be achieved either at the ideal situation of static synapses ($x = 1$), in which neither depression nor facilitation exist, and where all the neurotransmitters are available after a presynaptic spike or with dynamical synapses, in a balanced depression-facilitation situation ($x \approx 1$). Therefore, dynamical synapses could work equally efficient as the ideal situation (static synapses)
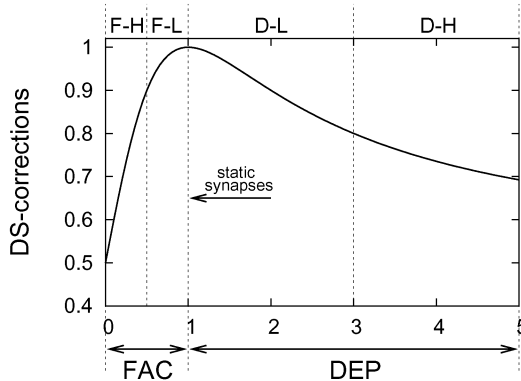
FIGURE 3. Energy discriminator in presence of dynamical synapses (*DS*). After dynamical synapses, the maximum storage corresponds with $P_{max} = N^2 \Psi(x)$. For static synapses, we have $P_{max} = N^2$ [1]. Therefore, the function $\Psi(x)$ is interpreted by the corrections to the static synapses situation due to short-term plasticity. These corrections, e.g. $\Psi(x)$, are represented in this figure as a function of x, the relative degree of change among depression versus facilitation after stimulus presentation. As explained in the main text, the variable x defines the nature of the synapse, and has mainly four different phases. $x > 1$ and $x < 1$ denotes respectively depressing (*D*) or facilitating (*F*)synapses. In addition, we distinguish for both depression and facilitation the situations of high (*H*) or low (*L*) limit. Balanced depression-facilitation synapses ($x \approx 1$) could have the maximum storage, the same as the ideal-unrealistic situation of static synapses.

regarding storage.

In summary, we have studied the familiarity discrimination in the framework of attractor neural networks, allowing to compare computer simulations with some theoretical calculations. We have introduced an alternative familiarity measure, the time derivative of the energy, that depends on the dynamics of the activity, and therefore is influenced by time evolution and by random fluctuations. In addition, we have studied how short-term synaptic plasticity affects familiarity discrimination. Although in general, maximum storage with familiarity discrimination depends on the details of the considered activity dependent mechanisms, as a result, the balanced depression-facilitation scenario, could store the maximum number of patterns (identical to the case of static synapses).

ACKNOWLEDGMENTS

REFERENCES

1. R. Bogacz, and M. Brown, *Hippocampus* **13**, 494–524 (2003).
2. A. Yonelinas, *J. Mem. Lang.* **46**, 441–517 (2002).
3. J. Hopfield, *Proc. Natl. Acad. Sci. USA* **79**, 2554–2558 (1982).
4. M. Brown, and J. Aggleton, *Nat. Rev. Neurosci.* **2**, 51–62 (2001).
5. L. Standing, *Q. J. Exp. Psychol.* **25**, 207–222 (1973).
6. M. Brown, F. Wilson, and I. Riches, *Brain Res.* **409**, 158–162 (1987).
7. M. Brown, and J. Xiang, *Prog. Neurobiol.* **55**, 149–189 (1998).
8. M. Rugg, and A. Yonelinas, *Trends Cogn. Sci.* **7**, 313–319 (2003).
9. D. Amit, *Modeling brain function: The world of attractor neural networks*, Cambridge University Press, 1989.
10. J. Cortes, M. van Rossum, and A. Greve, *In preparation* (2006).
11. A. Robins, and S. McCallum, *Neural Netw.* **17**, 313–326 (2004).
12. L. Abbott, and W. Regehr, *Nature* **431**, 796–803 (2004).
13. M. Tsodyks, K. Pawelzik, and H. Markram, *Neural Comp.* **10**, 821?835 (1998).
14. J. Torres, J. Cortes, J. Marro, and H. Kappen, *Neural Comp. In press. q-bio.NC/0604019* (2006).
15. H. Markram, Y. Wang, and M. Tsodyks, *Proc. Natl. Acad. Sci. USA* **95**, 5323–5328 (1998).