Journal of Computational Neuroscience 13, 187–205, 2002

The Mystery of Structure and Function of Sensory Processing Areas of the Neocortex: A Resolution

ANDRÁS LŐRINCZ, BOTOND SZATMÁRY AND GÁBOR SZIRTES
Department of Information Systems, Eötvös Loránd University, Pázmány Péter sétány 1/C, Budapest, Hungary H-1117
lorincz@inf.elte.hu
botond@inf.elte.hu
guminyul@ludens.elte.hu

Received September 25, 2001; Revised May 28, 2002; Accepted May 28, 2002

Action Editor: Alessandro Treves

Abstract. Many different neural models have been proposed to account for major characteristics of the memory phenomenon family in primates. However, in spite of the large body of neurophysiological, anatomical and behavioral data, there is no direct evidence for supporting one model while falsifying the others. And yet, we can discriminate models based on their complexity and/or their predictive power. In this paper we present a computational framework with our basic assumption that neural information processing is performed by generative networks. A complex architecture is 'derived' by using information-theoretic principles. We find that our approach seems to uncover possible relations among the functional memory units (declarative and implicit memory) and the process of information encoding in primates. The architecture can also be related to the entorhinal-hippocampal loop. An effort is made to form a prototype of this computational architecture and to map it onto the functional units of the neocortex. This mapping leads us to claim that one may gain a better understanding by considering that anatomical and functional layers of the cortex differ. Philosophical consequences regarding the *homunculus fallacy* are also considered.

Keywords: auto-associator, information processing, entorhinal-hippocampal loop, information maximization, neocortex

1. Introduction

In the last decade numerous studies and models dealt with the properties and the possible functions of the hippocampus and its environment. This interest is a consequence of the ever increasing number of new experimental findings, which highlight the striking complexity of the memory. However, due to this complexity, the most of the models are assigned to describe some particular features and areas of the hippocampal(HC)-neocortical(NC) memory circuits. (There are several excellent reviews, like Squire (1992b), or Hasselmo and McClelland (1999) and O'Reilly and Rudy (1999).) Although, to give a complete overview is beyond the scope of this paper, it is worth mentioning some models, which have proved useful in forming ours. The majority of the models have been developed to describe one part (mainly the CA3 field) of the hippocampus, because of the exciting freedom of interpretation of its recurrent structure (see, e.g. Levy, 1996). Besides these specific models, new integrating models of the HC have been also presented (Rolls, 1989; Hasselmo et al., 1996; Lisman, 1999; Eichenbaum, 2000) (see also the collection of theoretical papers and

the references therein (Ed.: M.A. Gluck, 1996)). It is known though, that the hippocampus is deeply embedded in the neocortical information flow through the entorhinal cortex (EC). This fact explains the emergence of a few EC-HC models like McClelland et al. (1995), Myers et al. (1995), Lőrincz (1998), Lőrincz and Gy. Buzsáki (2000) and Rolls (2000). The model in McClelland et al. (1995) emphasizes the necessity of a dual system for the seemingly contradictory tasks of learning specifications and allowing for generalization.

From the computational point of view, we should mention the model of Gluck and Myers (1993), which was designed to perform reconstruction *and* classification together for modeling some properties of the hippocampus. The reconstruction idea gains also importance in a recent model of Steinvas et al. (2000), in which reconstruction is used as regularization constraint in classification task for creating better representations. Reconstruction will play a central role in our model, too. As the earliest neurophysiological experiments focused on vision almost exclusively, there have been many models dealing with the structure and possible functions of the neocortical layers, especially within the LGN and the V1 area. As the complexity increases, the number of integrating models decreases. For example, Rao and Ballard (1999) have recently proposed an integrating model of the neocortical information processing. They explored a Kalman-filter analogy to cope with the input and system uncertainties (treated as noise) and presented a hierarchy for error correction and prediction using top-down inference from higher levels. Although the Kalman-filter idea—borrowed from control theory—could be an efficient and plausible function for hierarchical sensory processing, the mapping of the proposed function onto the anatomical, neurophysiological findings has not been fully elaborated. For spatial learning tasks, another Kalman-filter based model has been proposed with a biological mapping to the hippocampus (Bousquet et al., 1999).

In this paper we present a hierarchical structure for modeling the organization of long term memory. Although our model shares many similarities with previously proposed ones, its construction is different, because it is built on a single theory of generative networks. We claim that all structure components are essential consequences of the mathematical theory that underpins our model. We also argue that the aforementioned mapping to the neurobiological substrates can be advanced to the neocortical areas as well. The price to pay is that this extension is meaningful only if we drop the traditional association between anatomical structure and computational function. Our approach can be seen in many respects as a unifying view. For example, while most theoretical models can be classified by their feedforward (Hubel and Wiesel, 1962; Miller, 1994; Douglas et al., 1995; Suarez et al., 1995) or feedback structure (Somers et al., 1995; Ben-Yishai et al., 1995; Troyer et al., 1998; Adorján et al., 1999; Miller et al., 1999; Rao and Ballard, 1999; Koch and Poggio, 1999; Riesenhuber and Poggio, 1999), our work unites these different approaches without sacrificing their main advantages.

The paper is constructed as follows. Results are presented in Section 2. First, in this section we present a generative network with a reconstructing loop structure. The architecture is based on the general hypothesis on the use of representations. To demonstrate the proposed model's capacity and to display the working of the structure, some computer experiments are also presented. The hierarchical architecture is mapped to the neocortex in the discussion (Section 3). This section also includes philosophical aspects and describes our motivation behind the proposed model. For the sake of self-containment, a short overview of the mathematical details and the corresponding references are given separately in Appendix A (Section 5) and the detailed mapping and the supporting biological considerations are presented in Appendix B.

2. Results

2.1. Model Construction

2.1.1. The Core Model. In modeling neural systems three aspects should always be considered (Simoncelli and Olshausen, 2001): first, the task to perform, second, the definition of the environment and third, the set of biological constraints (like cell level neural behavior, connection systems, and so on). Aside from the third point we face the problem of designing a general information processing system. Our approach can be characterized as a 'grey-box' method (Oussar and Dreyfus, 2001). Contrary to the black-box methods, in which the goal is to give the best relation among the inputs and the outputs, while ignoring the 'inside' of the system, we first tried to describe the universal 'goal' of the neural systems. Our working hypothesis is as follows:

– there exists an internal representation of the external world, of the system itself and of their interaction.
– the input and the internal representation are mutually sensible, if the input itself can be re-generated in a top-down fashion by using the representation.

An architecture with this reconstruction or regeneration property is called generative network (Hinton and Ghahramani, 1997) and has been proposed, for example, as the underlying mechanism for perception by Horn (1977). The following equations describe the relation of the input, representation and the reconstructed input:

$$\mathbf{h} = \mathbf{W}\mathbf{x} \tag{1}$$

$$\hat{\mathbf{x}} = \mathbf{W}^{+}\mathbf{h} \tag{2}$$

where $\mathbf{x}$ is the external input, $\mathbf{h}$ is the internal representation, $\hat{\mathbf{x}}$ is the reconstructed input. $\mathbf{W}$ denotes the input-to-representation transformation (bottom-up, (BU) transformation) and $\mathbf{W}^{+}$ stands for the (pseudo) inverse of $\mathbf{W}$. Figure 1A shows a schematic draw of this reconstruction network.

The representation of the internal representation is the reconstructed (internally generated) input. Note that such reconstruction is also an auto-association: The input has an associated output which is similar to the input. It is worth noting that Hopfield networks (which are often used e.g. to simulate one part of the hippocampus (Káli and Dayan, 1998, 1999)) with dynamic attractors, for example, may represent a *single layer* auto-associator but without a separate layer for internal representation. That is, autoassociation does not naturally involve reconstruction. Reconstruction networks with internal representation units have been studied for signal compression in the early eighties (see e.g., the seminal work of Hinton and Sejnowski (1983)).

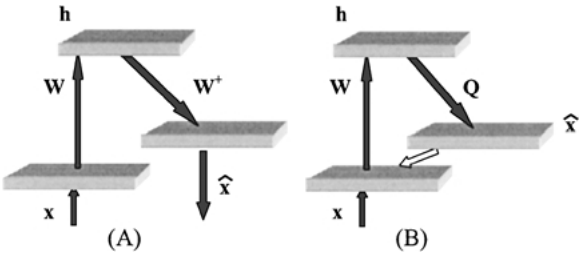

Figure 1. Basic reconstruction networks. (**A**) Simplest: reconstruction network: bottom-up (BU) and top-down (TD) transformations invert each other. (**B**) BU and TD transformation do not invert each other: BU error correction is used to correct the internal representation. Notations: The grey layers contain the computational units, black arrows denote excitatory connections, the white arrows are for the inhibitory connections $\mathbf{x}$: input, $\hat{\mathbf{x}}$: reconstructed input, $\mathbf{W}$: BU transformation, superscript $+$: (pseudo) inverse operation, $\mathbf{Q}$: TD transformation, $\mathbf{h}$: model layer activity.

Reconstruction network of Fig. 1A, however, is constrained: The input-to-internal representation transformation (which will be called bottom-up (BU) transformation) and internal representation-to-reconstructed input transformation (top-down (TD) transformation) should invert each other. Unfortunately, this constraint and feedforward working *together* do not fit the local property of Hebbian learning (which is thought to be the major source of neural plasticity) and poses a serious challenge for a neurobiologically plausible model.

2.1.2. Model Extension. As we have seen, inverting connection structures are not feasible in a neural system. As a consequence, improper (non-inverting) BU and TD ($\mathbf{W}$ and $\mathbf{Q}(\neq\mathbf{W})$, respectively) transformations produce error in reconstruction. This error can be processed in a bottom-up fashion to correct the internal representation. Correction is added to the internal representation and involves summation of the BU signals in time, i.e., temporal integration (on the internal time scale of the loop). The next equation describes the dynamic evolution of the internal representation:

$$\Delta\mathbf{h} \propto \mathbf{W}(\mathbf{x} - \mathbf{Q}\mathbf{h}). \tag{3}$$

In other words, the temporal integration of the computed error should take place at the internal representation level, e.g., by a self-excitatory recurrent collateral system.[1] That is, the aforementioned transformations should be complemented with internal dynamics as shown in Fig. 2. In mathematical terms, self-excitation, denoted by $\mathbf{M}$ in Fig. 2B, corresponds to recurrent identity transformation (the simplest case). Activity of the model layer is perfect, if it is able to reconstruct the input. Let us add this term to the model equation:

$$\Delta\mathbf{h} = \mathbf{W}(\mathbf{x} - \mathbf{Q}\mathbf{h}) + \mathbf{W}(\mathbf{x} - \mathbf{Q}\mathbf{h}). * \mathbf{M}\mathbf{h}, \tag{4}$$

where '(.*)' stands for component-wise multiplication. In line with our starting hypothesis, the equation describes the iteration that gradually improves the internal representation in order to reconstruct the input. It can be seen that the difference between the input and its reconstruction (i.e., the reconstruction error) has the

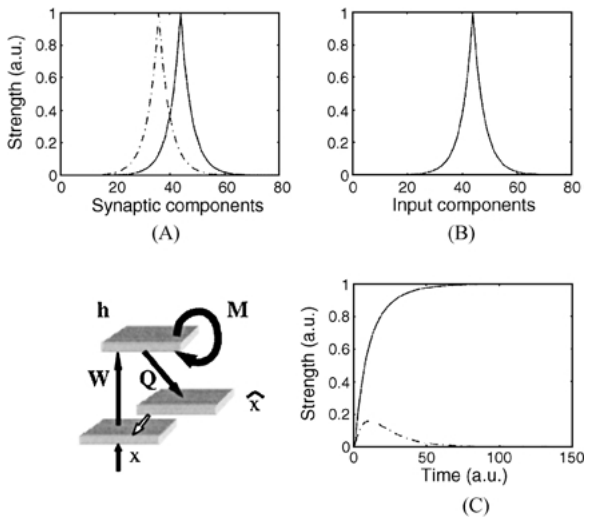

Figure 2. Dynamics of reconstruction network. (**A**) Synaptic component values in a two column TD matrix. The values are connected for the sake of better visualization. BU and TD matrices are transposed of each other. (**B**) Input components. (**C**) Time evolution of hidden layer activities. Lower left corner: Corresponding architecture. Notations: The recurrent collaterals of matrix **M** form identity matrix and execute temporal integration. Other letters are the same as on Fig. 1.

most fundamental role in governing the activity change, because both parts of the right hand side include this error.[2] The reconstruction error in Eq. (4) is almost exclusively determined by the input and the TD connections (denoted by **Q**). For example, if $\mathbf{W} = \mathbf{Q}^T$, where superscript T denotes transposition, then Eq. (4) minimizes the cost function

$$J = \frac{1}{2}(\mathbf{x} - \mathbf{Q}\mathbf{h})^T(\mathbf{x} - \mathbf{Q}\mathbf{h}), \tag{5}$$

which is the dot product of the reconstruction error with itself. In turn, after relaxation, the effect of Eq. (4) is as if the input **x** would be multiplied by the pseudoinverse of matrix **Q**.[3] If $\mathbf{W} \neq \mathbf{Q}^T$ then minimization is modulated by the BU matrix. When this BU modulation can be neglected (e.g., when matrix **Q** is invertible) the internal representation depends only on the TD connections and the input. If reconstruction is perfect, the TD connections are exclusively responsible for the 'success' of the reconstruction. It may be worth mentioning here that for matrix **Q**, Hebbian learning between representation and reconstruction error (i.e., $\Delta\mathbf{Q} \propto (\mathbf{x} - \mathbf{Q}\mathbf{h})\mathbf{h}^T$) minimizes the cost function of Eq. (5) and tunes **Q** to invert the rest of the loop. As a consequence, when learning of matrix **Q** stops, the TD matrix inverts the rest of the loop. We say that from the point of view of forming the internal representation, the *TD connections contain the learnt knowledge*. For this reason, we shall identify the TD transformation with the long-term memory (LTM) of the system. Other transformations within the loop may not constitute an organic part of the LTM in this model. On the other hand, for every reconstruction network, the input of the network can be regarded as perceived (sensory) input, so the BU process will be called sensory processing. By this labeling, we can distinguish unambiguously the information flows in the loop.

Finally we note that the time scales of the external world and the system are different, so temporal integration, while resulting in temporally convolved signals, seriously threatens the generative functioning and may produce temporal inadequacy.

2.1.3. Immediate Consequences. We list some consequences of the reconstruction idea:

1. Noise, e.g., noise of the perceived input does not contain information → noise should be filtered out, 'denoising' of input signal is necessary. Local denoising is necessary because the communication is noisy, too.
2. In order to make the signal reconstruction as perfect as possible → BU transfer of (denoised) information should be maximized (for a review on information maximization methods, see, e.g. Hyvärinen (1999b)) (that is information loss should be minimized).
3. Missing input components (e.g., if occlusion takes place) should be substituted or added to the input → correlation based pattern completion is also essential.

The second and the third requirements, however, pose the following problem: Maximization of BU information transfer can be accomplished by *minimizing mutual information* (MMI) amongst output components (Cover and Thomas, 1991; Bell and Sejnowski, 1995). MMI requires, that correlations of all orders need to be removed. MMI can be performed by different versions of independent component analysis (ICA, or, blind source separation (Jutten and Herault, 1991; Comon, 1994; Laheld and Cardoso, 1994; Bell and Sejnowski, 1995; Amari et al., 1996; Karhunen et al., 1997; Amari and Nagaoka, 2000), for a review, see Hyvärinen (1999b)). Elimination of higher order correlations, on the other hand, counteracts pattern

completion (3), because it heavily uses those correlations. In turn, it is impossible or at least inefficient to do pattern completion for components with minimized mutual information (MMI). And yet, MMI is most attractive as a first approximation, because it can be used for *local* probability estimation and denoising, besides the optimization of information transfer. Therefore, while preserving MMI, we introduce another layer that facilitates pattern completion and builds on the denoised MMI outputs.

2.1.4. Making Sense of Incomplete Patterns. Pattern completion can be considered as searching and adding missing sub-components to the output. There have been excellent theoretical developments along this line. Based on the assumption that the presence or absence of components is important in human signal processing, Charles and Fyfe (1998) have developed positive coding networks. Their model was clearly inspired by neurobiology (e.g., using only localized information, and assuming that synaptic connections cannot change sign). The proposed network is restricted to positive outputs. Others focus on discovering positive sub-components in the internal representation. The connections and the representations are both forced to be positive. This approach, initiated by the seminal experiments of Biederman (1987) on recognition by components, is called non-negative matrix factorization (NMF) (Paatero and Tapper, 1994; Lee and Seung, 1999, 2001).[4] We adopt this latter approach in our model and introduce an additional layer with positive coding properties. Note that temporal prediction can be seen as time-directed temporal pattern completion. Furthermore, temporal integration of a differential equation, such as the reconstruction dynamics, is also a special kind of temporal pattern completion. Thus, the new (spatio-temporal) pattern completion layer that follows the BU denoising stage may also be responsible for temporal integration. In this way the constraint of identity transformation imposed on the recurrent collateral system is resolved. In other words, **M** in Eq. (4) may differ from the identity matrix.

In summary, BU transformation works on the reconstruction error and the resulting output signal is directed to another layer, which

– facilitates the discovery of sub-components in the input
– executes spatio-temporal pattern completion
– originates TD reconstruction.

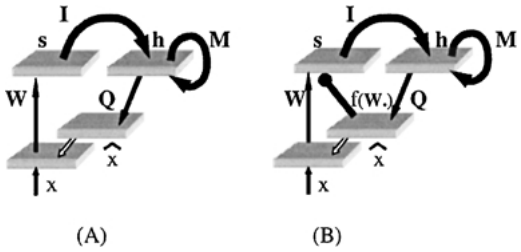

Figure 3. Extended reconstruction networks. (**A**) BU information can be completed. (**B**) BU information can be completed and denoised. The arrow with circle head denote non-linear transformation. **x**: input, $\hat{\mathbf{x}}$: reconstructed input, **W**: BU transformation, **Q**: TD transformation, **s**: BU processed reconstruction error, **h**: model layer activity with spatio-temporal pattern completion abilities. The dot in expression $f(\mathbf{W}.)$ denotes the vector that matrix W and then the non-linear transformation $f(.)$ acts upon. Matrix **W** within the brackets of the non-linear transformation $f(.)$ is approximately the same as it is in the outer loop (For details, see Appendix A).

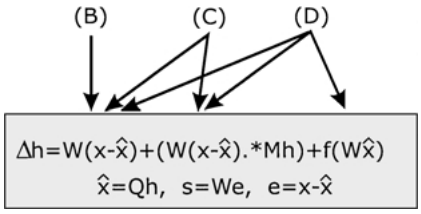

Figure 4. The governing equation of hidden layer activities. The basic equation that defines the activity state of the internal representation. The notions are as in Figs. 1–3. The capitals in parenthesis correspond to subfigures and denote the essential parts of the right hand side of the equation that describes the working of the corresponding system. (**B**) Fig. 2, (**C**) Fig. 3A, (**D**) Fig. 3B. In doing so the gradual extension of the base model can be easily understood. For example, arrows from '(C)' indicate that model C can be defined by only the two first expressions. One special notation is used: '(.*)' stands for componentwise multiplication. **s,** **e**, and $\hat{\mathbf{x}}$ refer to notations of Figs. 1–3.

From now on we refer to this latter layer as the model (or hidden) layer of the input signals. The corresponding architecture is depicted on Fig. 3C. Components of the reconstruction equations are summarized in Fig. 4.

2.1.5. Denoising the Input. To meet requirement 1, the system should exclude meaningless (e.g., measurement or sensory) noise from processing. As recent theoretical findings show, MMI can be easily combined with noise filtering (Olshausen and Field, 1996; Olshausen, 1996; Hyvärinen et al., 1999; Hyvärinen, 1999a). The essence of the proposed ideas can be summarized as follows:

1. In general, the MMI representation could assume super-Gaussian or sub-Gaussian distribution. For natural data (images, sounds, etc.), maximization of information transfer typically gives rise to sparse representations (i.e., super-Gaussian distribution) (See e.g., Field, 1987; Olshausen and Field, 1996).
2. The unstructured part of the input has maximum entropy, i.e., it contains no information and needs to be removed.
3. Noise filtering of MMI components is approximately equivalent with local thresholding.
4. Thresholding of each components of the MMI processed *reconstruction error* is local, too. Thresholding of each component of the MMI processed reconstruction error depends on the amplitude of the corresponding MMI components and not on the amplitude of the MMI processed error.

MMI components, however, are not explicitly available in our model: only the BU processed error of the MMI components is present. Nevertheless, the single role of these MMI components is to help the thresholding, which can be achieved by a transformation starting from the reconstructed input followed by the thresholding. Interestingly, the transformation—apart from the non-linear thresholding—is approximately equal to the transformation produced by the BU process (See Appendix A). Noise filtering of any MMI components can be accomplished by transforming the reconstructed input. If the transformed signals are above the threshold, then they can open the corresponding MMI error channels. In turn, the bottom up channel has to be extended with a non-linear gating functionality. The corresponding extension of the architecture is shown in Fig. 3D.

2.1.6. Order of Learning. So far we described the information flow in a specific generative network, in which different functions, like separation, denoising or pattern completion have been assigned to different computational units (layers). This section is intended to show the self-organized tuning of the loop induced by external inputs. Working and learning proceed as follows: In a well tuned system the TD matrix inverts the rest of the loop. Thus, any training input is reconstructed in one cycle. The internal representation of familiar inputs is computed as a true two-step (three-layer) feedforward processing. For novel inputs, many of the MMI components can be non-zero and small. Such activity pattern is compromised, sometimes abolished by thresholding. So, novel information will not be conveyed completely to the internal representation layer. Instead, BU learning will take place to maximize information transfer. During maximization, the effect of thresholding diminishes and the loop will start to process the new information.[5] At this stage learning of positive coding of the novel information can get started. The appearing reconstruction error will push the model layer activity towards its correct value. The Hebbian 'delta-rule' encodes the information into the TD transformation, that is it forms the inverse transformation of the rest of the loop. After learning, the forming of the internal representation will be as fast as in non-recurrent feedforward networks.

2.2. The Other Half of the 'Grey Box': Mapping to the EC-HC Architecture

This section 'fills in' the missing details and present the EC-HC loop model by taking into consideration the properties of the biological system to be modeled (See the third aspect in Simoncelli and Olshausen (2001)).

As it was emphasized earlier, we cannot avoid facing the problem of convolution. It is feasible to cope with it *within* the BU processing channel. The mathematical procedure, which can do this job is called blind source deconvolution (BSD). BSD, in general, is very demanding in terms of neuronal units and the level of connectivity amongst those units (Lee et al., 1997). Leaving the theory aside for a moment, if we look at the brain's anatomy and its known functional relations, it turns out that loopy corrections may occur at many different level of the hierarchy, which compromise the forming of internal representation in a pure feedforward manner at higher levels. These loops in our model produce exponential temporal smoothing (temporal convolution). It seems that the effect of exponential temporal smoothing may be a general feature: This effect also exists at the level of the LGN, for example. It has been reported that LGN efferents exhibit temporal smoothing with a kernel function of an approximately exponential envelope for longer times (see, e.g., Saul and Humphrey, 1990; Cai et al., 1997; Wimbauer et al., 1997; and references therein). Fortunately, BSD of exponential kernels is much less demanding and that the dentate gyrus *has* the optimal architecture and working to perform BSD of such exponentially convolved inputs (Lőrincz and Gy. Buzsáki, 2000). This structure (of DG) is unique, since it only exist in the hippocampus. The question that comes to mind is whether BSD

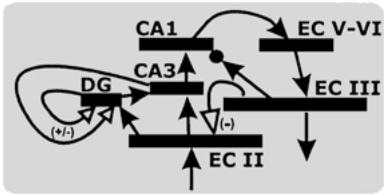

Figure 5. Hippocampal-entorhinal loop. Schematic drawing of major fields and connections in the entorhinal cortex (EC)-hippocampal formation and their functions according to our model. The hippocampus communicates with the perirhinal, parahippocampal and neocortical regions by way of the EC (from cortex; to cortex). The main information flow in the EC is from the deep (V-VI) to superficial (III and II) layers. According to our mapping, long-term memory (LTM) is stored in the excitatory synapses between deep and superficial layers of the EC and in analogous synapses of the neocortex (not shown). The EC directly innervates all major hippocampal fields. The difference between neocortical information and processed signal in the hippocampus-layer V-layer II loop (i.e., the reconstructed input) is regarded as the reconstruction error in the model. This error signal drives the dentate gyrus and the CA3 field. Granule cells of the fascia dentata and mossy cells of the hilus form delay lines and function primarily as a deconvolution network. The EC provides the input to the CA3 field (perforant path and mossy fiber synaptic matrices), and this input is whitened by the CA3 field. Further minimization of mutual information occurs in the CA1 field. Layer III of the EC is re-mapped onto the CA1 field and performs denoising. The resulting noise-free independent components are transformed to a positive coding network which can perform spatio-temporal pattern completion. Outputs of this positive coding network are used to train LTM in the EC.

is sufficient at the top level of sensory processing (performed by DG).

Though it seems a technical question, learning MMI components can be accomplished at different convergence speeds. It is known (Amari et al., 1996; Amari and Nagaoka, 2000) that forming of MMI components (that is BU maximization of information transfer) can be accelerated if the process is divided into two stages. The first one is so 'whitening' (decorrelating) stage, which precedes the actual MMI stage. In turn, BU processing may be further divided into two stages. Interestingly, if our mapping is correct, our brain does apply the two stage learning in the EC-HC loop. The full architecture with the additional whitening and BSD subsystems in BU processing is shown in Fig. 5. More detailed description of this loop is provided in Appendix B.

2.3. *Computational Results*

The capabilities of our model are demonstrated on a 'toy problem', in which the goal is to reconstruct bar[6] images (i) corrupted by noise, (ii) having missing parts and (iii) corrupted by noise and having missing parts. The inputs of the system are grey scale images of 256 components which can be arranged in a 16 by 16 pixel image. Each 'image' consists of 2 or more randomly chosen bars positioned horizontally and vertically. Some non-linearity is present in this problem; inputs for overlapping bar points are not summed. Therefore, independence of the sources is only approximate in this example. It may be worth noting that independence, however, is not necessary for noise filtering (Hyvärinen et al., 1999; Hyvärinen, 1999a). Zero mean Gaussian noise was added to all images. Noisy inputs were re-scaled to fall between zero and one. For investigating hierarchical processing, we have taken additional loops arranged in a topographic manner as well as a higher layer dealing with the processed information of those topographical lower layers. This double layer hierarchy was tested on the combined problem: The inputs were covered with noise and contained bars with missing components.

Learning was performed off-line. In this way some components of the EC-HC architecture can be neglected. The deconvolution architecture was left out and the two-step processing that speeds-up MMI formation was unified into a single transformation in the test phase, i.e., upon learning. The hierarchical setting of the simplified architecture is shown in the upper subfigure of Fig. 6. Simulation results for a single layer are shown in the lower subfigure of Fig. 6. Subfigure (A) and (B) depict one of the 'ideal' inputs and its noise covered variation, respectively. This latter (i.e., (B)) is a typical sample used for training and testing. Interestingly, the MMI process without the help of the NMF failed to reconstruct the input (subfigure (C)). The same was true when NMF has been applied only (subfigure (D)), probably because of the high noise content. At the same time the combined method 'predicts' higher pixel values at the places that correspond to the correct pixels of value 1 of the original (clean) input (see Fig. 6E).

Recall that input topography (that is the distance relation among the input components) has not been used by the algorithm. In turn, we have the freedom in interpreting the figures. One option is to think of the horizontal direction as different efferents from lower levels and the vertical direction as *time*. In other words, the algorithm may work on spatio-temporal inputs as it has been suggested by Lisman and Idiart (1995), Jensen and Lisman (1996), Lisman (1999) and has been also studied in Lőrincz et al. (2001b).

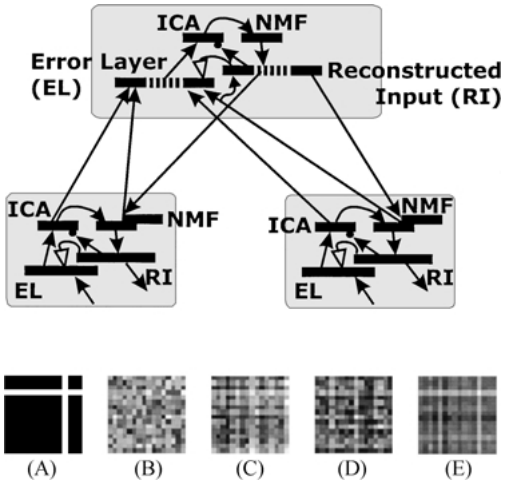

Figure 6. Hierarchical architecture, inputs and outputs for one layer. *Upper sub-figure*: Two-layer hierarchy. Lower layers provide inputs (their hidden variables) for higher layers. Reconstructed input of second layer is first a copy of the input of the second layer (copying is represented by the small curved arrows). Reconstructed input of the second layer is then formed by the loop of the second layer and it overruns hidden variables at first layers. Hidden variables are subject to NMF iteration. Dashed and dash-dotted grey arrows represent identity transformations. *Lower sub-figure*: (**A**) perfect input without noise, (**B**) noise covered input, (**C**) reconstructed input (RI) using only denoising, (**D**) RI using only NMF, and (**E**) RI with combined denoising and NMF. Note the improved reconstructions for the combined method compared to single denoising or single NMF algorithms.

The emerging basis sets for the three methods (MMI alone, NMF alone and the combined hierarchy) are shown in Fig. 7. According to the figure, NMF is not capable to overcome the high noise content, whereas MMI with denoising capacity discovers the high-order

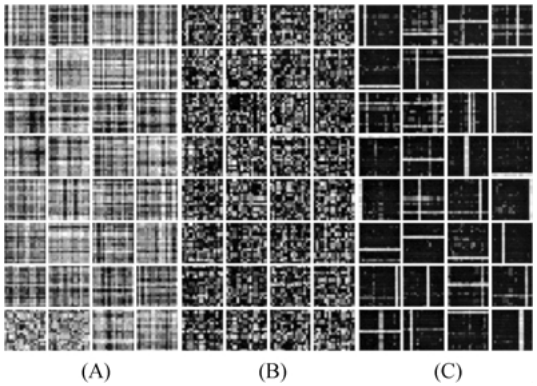

Figure 7. Single layer basis sets for MMI alone (**A**), NMF alone (**B**) and for the NMF using MMI outputs (**C**). Inputs have 256 ($=16\times16$) dimensions. Number of filters: 32 ($=2\times16$). For each method, all 32 filters are depicted.

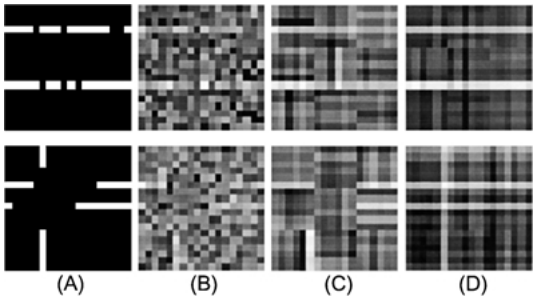

Figure 8. Improved noise filtering and pattern completion in the hierarchy. Sub-architectures of the first layer have $6\times6=36$ input dimension. Dimension of NMF hidden vectors of first layer units is 12. First layer is made of $3\times3=9$ units. Input of the second layer has 108 dimensions. Dimension of the hidden vector of the second layer is 36. *Upper row*: Pixels of the inputs are missing. *Bottom row*: Pixels and sub-components of the inputs are missing. (**A**) Original input with missing pixels and sub-components, (**B**) input to the architecture, (**C**) reconstructed input using first layer reconstructions only, (**D**) reconstructed input using the full hierarchy.

line-like correlations in the input but is not capable to decompose the input into sub-components. The combined method shows robustness against noise and is able to extract sub-components (single vertical and horizontal lines).

In the completion tasks we studied the efficiency of the combined method in a hierarchical structure (contrary to the noise study, when a single compartment model was used). We applied a two layer hierarchy with a layer consisting of several compartments as shown in the upper subfigure of Fig. 6. Each compartment worked on a localized sub-area of the input, and the second layer collected the results from the first layer over the entire input range (see Fig. 6). Two types of inputs were generated. The upper row of Fig. 8 shows an example of the original (noise-free) input with missing pixel components. The lower row depicts the case when full sub-components from lower layer are missing. Noise-free, incomplete inputs are shown in column (A). Visual examination reveals that input reconstruction improves both for the missing pixel case as well as for the missing subcomponent case.

3. Discussion

First, we would like to highlight some appealing properties of our proposed hierarchical structure. This architecture assumes two 'noise components', (i) a unstructured noise component, which should be removed and (ii) another type of noise, which is responsible for the

presence or the absence of sub-components of structures. The removal of the unstructured noise (strict denoising) and the recognition of the subcomponents take place at different stages in our architecture. The two-stage procedure can be performed by strictly local computations using local filtering and learning rules (see Lőrincz and Gy. Buzsáki (2000) and the Appendix A).

3.1. *Mapping to the Neocortex*

In the previous sections we have constructed a hierarchical information processing system and mapped it onto the EC-HC loop. This region plays many specific roles in the brain, but its loopy connection system is not unique: feedback and loops seem to be present everywhere. Based on this universality we try to map the model onto the neocortex, too. We enumerate the relevant anatomical and physiological data and assign the model tasks to different subregions. The neocortex is made of six sub-layers as it is shown in Fig. 9. The figure depicts the most prominent connections between these sub-layers. Input typically arrives to layer IV. Layer IV neurons send messages to layer II and layer III. Furthermore, layer IV neurons send messages also to layer VI. Superficial neurons provide some output down to layer V and VI. There are connections between layers V and VI. Layer V provide feedback to layers II and III. (For a review, see, e.g., Callaway (2000).) The main output to higher cortical layers emerges from layers II and III. The main feedback to lower layers and subcortical structures is provided by layer V. This structures seems to be reversed when compared with our architecture.

The discrepancy between theory and the anatomical structure may be resolved by questioning the identification of functional and anatomical layers. Dropping this 'preconception' and allowing for differing anatomical and functional layers the mapping can be made more

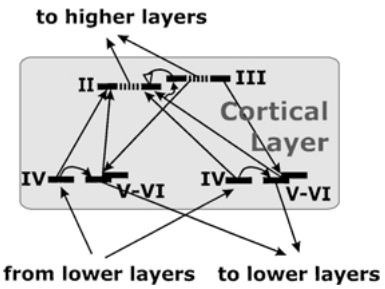

Figure 9. Neocortical layer.

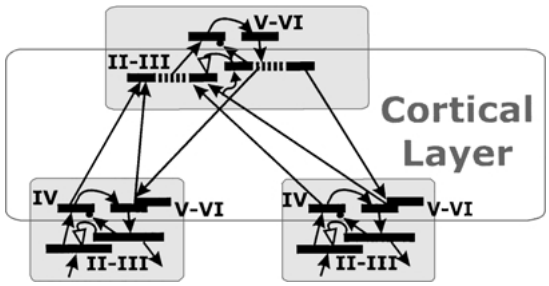

Figure 10. Neocortical hierarchy. For comparison, see Fig. 6.

plausible This view is depicted in Fig. 10. According to the figure, superficial layers of the lower anatomical layer *and* deep layers of the higher anatomical layer form *one* functional layer. Reconstructed input of second layer is first a copy of the input of the second layer (copying is represented by the small curved arrows). Reconstructed input of the second layer is then formed by the loop of the second layer and it overruns hidden variables at first layers.

As far as the EC-HC loop is considered, the loop is closed by the hippocampus, which is positioned on the top of the hierarchy as depicted in Fig. 11.[7] In other words, we hypothesize that many interacting loops coexist with a separate, slow-changing long term memory (which refer to their own experiences) and the entorhinal-hippocampal loop (with its own LTM), which organizes all loops' working to create and maintain the system's overall long term memory. In this view, the V1 performs two functions: layers V and VI of the primary visual cortex hold the internal representation of the LGN, while layers II and III represent the input and the reconstructed input of V2, respectively. At last we note that the above described hierarchy not only may offer a model of the long term memory, but

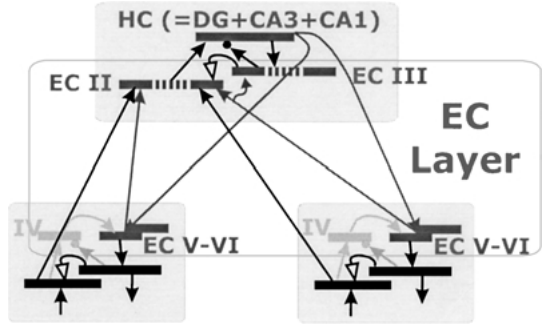

Figure 11. Hippocampus and its environment. For comparison, see Figs. 5, 9 and 6.

also provides a framework for further extension. This is essential, if the final goal is to provide a joint description of the different memory phenomena and their relation. The explorations toward this unified model has recently started: we proposed that the remaining transformations in the loop can be responsible for *implicit* learning phenomena. Low level perceptual priming can be explained by BU information maximization (Szirtes and Lőrincz, 2001), whereas category formation can be explained by training the associative connections of the internal representation layer (i.e., matrix **M**) (Aszalós et al., 1999; Sz. Kéri et al., 2002).

3.2. Philosophical Considerations: A Resolution of the Homunculus Fallacy

As it was described in the Introduction, we propose the generative network concept as the underlying idea of the neocortical hierarchy. This concept, however, can be seen as a consequence of a deeper philosophical idea. We feel that discussing this idea may be useful, even if it is a bit off the main stream.

Our thoughts are grounded on the hypothesis that representations do exist in brain (See e.g. the debates about the Representational Theory of Mind and its modern extension, the Computational Theory of Mind (Fodor, 1981; Churchland, 1989), but also (Dennett, 1987)). Though the debate on the existence and form of any innate representations (e.g. Smolensky (1989) and Gelder (1995) but also Dennett (1988)) has not yet settled down, the use of representations can hardly be avoided in any computational modeling. Furthermore, the universal goal of *any* modeling is to label meaningfully the building blocks of the phenomenon to be modeled and define their relation that fundamentally influences the existence of the given phenomenon.

Generally speaking, the processing of signals that may convey information can be considered as a transformation into another form that still carries the whole amount or just a piece of the original information. The environment feeds the system with some inputs and the system output represents (a part of) the environment. Whilst most models address the problem of coding inputs and making efficient internal representation, we are more concerned about the fundamental problem of making sense of these representations. In our view, the central issue of making sense or *meaning* is to provide answers to questions like 'what does it mean?' in terms of our past experiences, or 'how is it related?' in terms of known facts. In other words, making sense is inherently related to declarative memory. As a consequence, the homunculus fallacy (see e.g., Searle, 1992)—that the internal representation is meaningless without an interpreter—is of central importance. This fallacy claims that all levels of abstraction require at least one further level to become the corresponding interpreter. Unfortunately, the interpretation—according to the fallacy—is just a new transformation and we are trapped in an endless regression. This problem could be more than a philosophical issue. We are afraid that any model of declarative memory, or a model of structures playing role in the formation of declarative memory could be questioned by the kind of arguments provided by the fallacy.[8]

Our standpoint is that the paradox stems from vaguely described procedure of 'making sense'. The fallacy arises by saying that the internal representation should make sense. To the best of our knowledge, this formulation of the fallacy has not been questioned except in our previous works (Lőrincz, 1997c; Lőrincz et al., 2001a). One can turn the fallacy upside down by changing the roles: Not the internal representation but the *input* should make sense. Our proposal is that the *input makes sense* if the same (or similar) inputs have been experienced before and if the input can be derived or regenerated by means of the internal representation (Lőrincz, 1997c; Lőrincz et al., 2001a). All in all, the goal is to turn the infinite regression into a reconstructing loop structure and shortcut the fallacy. According to this approach the internal representation interprets the input by (re)constructing it. We have shown that the idea on generating the inputs from the internal representation may help to understand the structures thought to be responsible for declarative memory, such as the hippocampus and the adjacent medial temporal lobe structures.

3.3. Open Questions and Remarks

This paper was intended to give a consistent model of learning and memory forming. However, it is widely accepted that memory is not a single unit cognitive function. There are many implicit memory functions that may be intact even if most of the explicit memory functions are impaired. In our previous work (Szirtes and Lőrincz, 2001) we have shown that low level form of implicit memories (the so called perceptual priming) can be easily embedded into the described model. In another work, findings on 9-dot categorization (Knowlton and Squire, 1993) for the case of Alzheimer patients

(Kéri et al., 1999) was explained by means of the recurrent collateral system of the model layer (Aszalós et al., 1999; Sz. Kéri et al., 2002). Additional work is required to investigate the interaction among these phenomena of our generative structure.

Another open issue is the presence or absence of other time separating (that is deconvolving) structures besides the DG. Since information from almost every areas of the neocortex converges to the HC, it may be possible that its deconvolving ability is one of the features, which make it a central formation. It is possible that the DG could be a uniquely efficient structure to compensate for corrupted temporal information. Nevertheless, the feedforward nature of generative networks upon proper tuning diminishes temporal convolutions, too. Our approach accords with other models that emphasize the central role of HC in all aspects of information processing in the brain. However, our approach also implies that both maximization of BU information transfer and diminishing of temporal convolution can be accomplished without fast two-stage natural gradient MMI (the tentative role of the CA3 and CA1 fields) and the fast temporal deconvolution (the tentative role of the DG). In turn, we suggest that the hippocampal-entorhinal loop may do the same computational task as sensory neocortical areas but at a higher speed. Our model can be seen as a prototype model, which describes the common computational tasks of the entorhinal-hippocampal loop and sensory neocortical areas. The described model has already gained some support: some model predictions on the DG has been confirmed recently (Henze et al., 2001).

An important shortcoming of our model is that we are not able to describe hierarchical consolidation of the memory, the organization of the neocortical '*database*'. We regard consolidation as the slower tuning of the less flexible connection systems in the neocortex, it generally takes place at a longer interval. This fact is articulated by the retrograde amnesia studies, in which hippocampal lesions result in different memory loss depending on the time between the injury and the questioned memory element. It is not clear, how to 'transfer' pieces of LTM from the fast analyzing architectures, to the appropriate areas of the neocortex.

It is important to emphasize that the strong independence assumption behind the ICA algorithm serves to approximate the joint probability distribution. The factorized approximation of the joint probability distribution is called *mean field approximation* (see, e.g., Chandler, 1987). Mean field approximation has the advantage that probability estimation becomes local (i.e., component-wise) after ICA transformation has been done. In turn, this local estimation can be used to reject noise. In our architecture transformation and noise rejection are accomplished by two BU channels, which should contain almost identical demixing matrices. In one of these channels a gating non-linear function is also necessary for denoising.

The question that remains is whether the two BU channels can learn the same matrix or not. This question does not concern the learning of the ICA transformation, both channels can learn to separate. The question concerns the order (the index) of components, because ICA is invariant to the permutation of the components. For further considerations, the error processing BU channel will be called BUE channel, whereas *BUD* channel stands for the *denoising BU* channel. The joint output of the two channels is the BUE output.

From the point of view of Hebbian learning, we need to consider the input and the output of the BUD channel. The *input* to this channel is the reconstructed input, whereas the *output* of this channel is the BUE output, that is, the temporal derivative of the error of IRS layer. Tuning of any matrix performs time integration. In turn, apart from a constant factor and from the point of view of matrix tuning, the BUE output approximates the activities of the internal representation.

Another contribution to learning could be provided by accidental coincidences. Coincidences along one synapse may be generated by the complementary pathways in the loop structure, which are traversing through the internal representation to the appropriate component of the reconstructed input. The stronger these channels, the more likely the random coincidences will occur. This channel, by construction of our autoassociative loop, can be taken as approximately equal to the inverse of the denoising matrix. We shall argue below that this is a good approximation, indeed. In turn, the stronger the appropriate component of the transposed of the inverse matrix, the more likely that further learning—due to random coincidences—may occur. We note that accidental coincidences can be justified on the basis of that Ca^{2+}-permeable N methyl-d-aspartate (NMDA) channels or voltage-gated Ca^{2+} channels stay open for a relatively long time (see, Lőrincz and Gy. Buzsáki (2000) and references therein). Taking these accidental coincidences and subtracting the Hebbian contribution, the resulting learning rule is the ICA learning rule derived by Bell and Sejnowski (1995).

Interestingly, the reconstruction architecture makes the Bell and Sejnowski learning rule local. Given the ICA learning rule of Bell and Sejnowski (1995), the BUD matrix has the appropriate order of components, because—in a well tuned system—both BUD and BUE matrices invert the same TD matrix. Note however, that the Bell and Sejnowski learning rule is slower than the natural gradient learning rule (Amari et al., 1996) controlling the training of the BUE matrix (Lőrincz and Gy. Buzsáki, 2000). In turn, the BUD matrix can be different from the BUE matrix. According to the Bell-Sejnowski training rule, the BUD matrix follows the inverse of the TD matrix, whereas the TD matrix learns to invert the novel information processed by the BUE matrix. In turn, the order of learning and the consistence of the architecture is maintained in our model. We can also conclude that in our model denoising refers to the internal representation as it should do.

The recurrent architecture of the hippocampal-entorhinal loop was modeled first as an error correcting control system by Lőrincz (1997a, 1997b, 1998). An extended form of that model (Lőrincz and Gy. Buzsáki, 2000) served as a base to form a prototype system in the present work. Details of the hippocampal-entorhinal loop, according to our generative scheme (e.g., the CA3 the recurrent collateral system and the need for two-phase operation), are presented in those works. Similarities of this model family and other recurrent models based on control ideas (Bousquet et al., 1999; Rao and Ballard, 1999) are apparent. The major difference is that no assumption is made here that would favor a Kalman-filter. Instead of directly building a model to interpret experimental data, we intended to solve the homunculus fallacy and its conjugate problems, such as noise filtering and pattern completion. The resulting structure *can* be, e.g., a Kalman-filter.

Finally, we would like to note that the emerging duality between functional and anatomical areas/layers gives rise to several issues, which may question the model. One of the most intriguing issue for us is that the model provides no hint about the recurrent collateral system of the superficial layers of neocortical areas. It remains to be seen if this is a major deficiency of our model or not. In closing, a few final remarks are listed:

- Unstructured noise, which is filtered out at one stage may provide information for neurons with larger receptive fields in higher layers. In turn, reconstruction error should be further processed in a bottom-up fashion to discover spatio-temporal features of higher complexities.
- Although a hierarchical model is described here, the model does not explain the hierarchical (re-) organization of LTM. Computational results of the hierarchical architecture assumed pre-wired hierarchical topographical connections.
- TD connections from higher areas to lower areas are abundant but are not very active (see, e.g., Calvin (1999) and references therein). Our model suggests that in pattern completion problems the activities through these connections should increase.

Appendix A

A.1. Mathematical Framework

Since all the used algorithms in our model are well founded, this section is intended to give only a short summary. The interested reader may find the details in the references. Let $\mathbf{x} \in \mathbb{R}^n$ denote the input, where n is the dimension of inputs (e.g., pixels) and $\mathbf{x}$ is the input or the difference between the original and the reconstructed (approximate) inputs. Let's suppose the observed signals are made of linear combination of independent sources:

$$\mathbf{x} = \mathbf{A}\mathbf{s} + \eta, \tag{6}$$

where $\mathbf{A} \in \mathbb{R}^{nxm}$ is the mixture matrix, $\mathbf{s} \in \mathbf{R}^m$ is the unknown source set, and $\eta \in \mathbf{R}^n$ is the noise. In our case m is less than n. Parts of input with no structures (or with unrecognized structure) are treated as noise. The problem of finding the unknown sources and the separating matrix (that is approximating the inverse of the mixing matrix) is known as blind source separation (BSS) in signal processing literature and Independent Component Analysis in statistics, see Comon (1994) and Hyvärinen (1999b) and references therein. Based on this model we are looking for the sources, which problem is equivalent with minimizing the mutual information among the components:

$$\mathbf{s} \approx \mathbf{W}\mathbf{x}, \tag{7}$$

where W is approximately the pseudoinverse of the mixing matrix ($\mathbf{W} \approx \mathbf{A}^+$) We applied the FastICA algorithm of Hyvärinen and Oja (1997). Although the algorithm can be implemented in online fashion we used it in batch mode. It has been also shown that the

tuning of the weight matrix (the connection matrix) can be based exclusively on local information, that is the learning is Hebbian. Because most of the algorithms require preprocessing in solving BSS problem, we also whitened the data before further processing. Whitening can be considered as calculating the so called principle coordinates of the given data set (that is finding the directions, in which the variance is the largest) and then project and normalize the data on a unit circle spanned by these coordinates.

$$\mathbf{x}_w = \mathbf{W}^+\mathbf{W}\mathbf{x}, \tag{8}$$

where $\mathbf{W}^+\mathbf{W} \in \mathbf{R}^{nxn}$ is the projection matrix. It has been shown that this processing may be tuned by Hebbian learning rules (Lőrincz and Gy. Buzsáki, 2000). The projection eliminates second order correlations, too. The rest of the BSS transformation concerns the rotation of the coordinate axis to minimize mutual information.

In our computer experiments, the sparse code shrinkage (SCS, or denoising) algorithm of Hyvärinen et al. (1999) was used for training the denoising matrix: Let us denote the matrix of this denoising channel (SCS filter) with $\mathbf{W}^{scs}$. The tuning equation of $\mathbf{W}^{scs}$ is as follows:

$$\mathbf{W}^{scs}(k+1) = \mathbf{W}^{scs}(k) + \lambda(k)q(\mathbf{W}^{scs}(k)\mathbf{x})\mathbf{x}^T + \frac{1}{2}(\mathbf{I} - \mathbf{W}^{scs}(k)\mathbf{W}^{scs}(k)^T)\mathbf{W}^{scs}(k), \tag{9}$$

where k is the number of the training session, $\lambda(k)$ is the learning rate, and $q(.)$ denotes a non-linear function of the following form: $q(u) = -u\exp(-u^2/d^2)$. Here, $q(u)$ is applied separately to each component of the vector argument. Assuming that $\lambda(k)$ satisfies the Robbins-Monro condition, this set of equations has only one parameter, the kernel width, d.

In forming the hidden representation the following approximation was made. The output of the BU channel was multiplied by the pseudo-inverse of the ICA separating matrix. This projected output was then analyzed by the non negative matrix factorization algorithm. This inversion was intended to make possible fair comparisons among different methods and their combinations.

In NMF, the basic problem is the following: Let $\mathbf{X}^{dn}$ denote the matrix of the denoised inputs provided by the BU stage. Our goal is to factorize this matrix in the next form:

$$\mathbf{X}^{dn} \approx \mathbf{QH}, \tag{10}$$

where $\mathbf{Q}$ is the NMF basis set and $\mathbf{H}$ is the matrix of the hidden representations. Both matrices are constrained to have non-negative elements. It can be shown, however, that this problem can be only solved by iteration. The so called multiplicative update rules (a dual iteration) were used to minimize the Euclidean distance based cost function (Lee and Seung, 1999, 2001). The next iterative equations were used in batch mode for learning:

$$\mathbf{H}_{ij}(k+1) = \mathbf{H}_{ij}(k)\frac{(\mathbf{Q}^T(k)\mathbf{V})_{ij}}{(\mathbf{Q}^T(k)\mathbf{Q}(k)\mathbf{H}(k))_{ij}} \tag{11}$$

$$\mathbf{Q}_{ij}(k+1) = \mathbf{Q}_{ij}(k)\frac{(\mathbf{V}\mathbf{H}^T(k))_{ij}}{(\mathbf{Q}(k)\mathbf{H}(k)\mathbf{H}^T(k))_{ij}}, \tag{12}$$

where columns of matrix $\mathbf{H}$ represent the hidden representation vectors, $i = 1, \ldots, n$ (n is the dimension of the hidden vector), $j = 1, \ldots, m$ (m is the number of inputs). Matrix $\mathbf{H}$ is formed by Eq. (11). This iteration makes use of the pseudoinverse $\mathbf{W}^+$ of $\mathbf{W}$: $\mathbf{X}_j^{dn} = \mathbf{W}^+\mathbf{S}_j$, where $\mathbf{X}_j^{dn}$ and $\mathbf{S}_j$ denote the jth input and the jth output of the *linear* SCS transformation during the batch learning, respectively. In the learning phase both Eqs. (11) and (12) are used for the NMF iteration. In the working phase new inputs are generated and only the first equation is used for single vectors (the NMF matrix $\mathbf{Q}$ is not modified). It can be easily shown that the learning rule can be reformulated if on-line tuning is required. Let $\mathbf{x}^{dn}(t)$ denote the denoised input at time t. In this case the following learning algorithm is provided by differentiating the $\|\mathbf{x}^{dn}(t) - \mathbf{Q}\mathbf{h}(t)\|^2$ error term w.r.t $\mathbf{Q}$ and $\mathbf{h}$:

1. $t = 1$;
2. $\mathbf{h}(t) = \mathbf{h}(t) + \alpha_1\mathbf{Q}^T(\mathbf{Q}\cdot\mathbf{h}(t) - \mathbf{x}^{dn}(t))$
3. $\mathbf{h}(t) = \oplus(\mathbf{h}(t))$
4. $\mathbf{Q} = \mathbf{Q} + \alpha_2(\mathbf{Q}\cdot\mathbf{h}(t) - \mathbf{x}^{dn}(t)\mathbf{h}(t)^T)$
5. $\mathbf{Q} = \oplus(\mathbf{Q})$
6. if not converged, go back to (2), otherwise continue
7. $t = t + 1$, back to (2)

The function denoted by $\oplus$ filters out (substitutes with zeros) the negative components of its argument, α_1 and α_2 are learning parameters.

Appendix B

B.1. Collection of Biological Data That Supports Our Mapping onto the HC-EC Network

This section is intended to enumerate and describe some neurobiological data that are relevant for understanding the suggested architecture.

At the synaptic level, two phenomena are of central importance in supporting our model. For details, see Lőrincz and Gy. Buzsáki (2000). Model assumptions concerning formation of independent component include

- synaptic modification by large postsynaptic membrane depolarization of sufficient duration to open Ca^{2+}-permeable N-methyl-d-aspartate (NMDA) channels or voltage-gated Ca^{2+} channels (Bliss and Collingridge, 1993; Markram et al., 1997) allowing for 'effective coincidences'
- synaptic normalization mechanisms have been recently reported (Markram et al., 1997; Debanne et al., 1998; Bi and Poo, 1999; Abbott and Song, 1999).

As we believe that HC processing produces statistically independent inputs, the process which results in this independence property, has to be manifested within the system. Fast learning of independent component analysis occurs if the procedure is divided to two subprocesses: whitening and separation. Now we assign these tasks to different parts of the loop.

B.2. The CA3 Field May Perform Whitening (Decorrelation of Inputs)

The CA3 field receives a direct and indirect (by way of the DG) from EC layer 2 of the EC (Amaral and Wittner, 1989; Lopes da Silva et al., 1990; Tamamaki and Nojyo, 1993). The model claims that the CA3 layer decorrelates (whitens) the EC-conveyed information that is its task would be to eliminate the second order correlations and normalize the resulting data. The whitening process takes place both during both the theta and the sharp wave (SPW) states. In the awake (theta) state inputs from the perforant path and mossy fibers are whitened, whereas during the sleep phase (SPW) the subject of the whitening process involves are the signals generated by the recurrent collateral matrix structure of the CA3 region. Tuning of the recurrent collateral synapses is important and will be discussed below.

B.3. The CA1 Field May Perform Separation

Separation corresponds to a rotation of the coordination axis. The CA3-CA1 projections (Schaffer collateral synapses) are very divergent, so those may be capable for that transformation. A single CA3 pyramidal cell may innervate as many as 30,000 CA1 neurons (Li et al., 1994). It was hypothesized in the model that separation of whitened information takes place in these synapses. Separation requires a two-step (nonlinear and linear) operation (Amari et al., 1996; Wang et al., 1995; Wang and Karhunen, 1996; Karhunen et al., 1997). During theta behavior, CA1 pyramidal cells discharge sparsely, a feature that was attributed to nonlinear thresholding. In the linear (SPW) phase, population firing rate is high (Csicsvari et al., 1999) and normalization of synaptic strengths occurs. Because the CA3 activities are already whitened, the two-step bigradient-like tuning rule (Wang et al., 1995) may perform separation. In turn, the complex requirements of the separating function are not contradicted by the corresponding architecture and its working.

B.4. The EC Computes the Reconstruction Error and Represents the Independent Components (ICA)

The EC is both the input stage and the output stage of the autoassociator/autoencoder hippocampal architecture. In addition, the EC provides information about the error of the autoassociation (i.e., the reconstruction error). Layer III represents the neocortical input to the EC. Layer II receives the same input as layer III, but it is also influenced by Layer III. In addition, the input is combined with the intracolumnar efferents of layers V-VI of the EC. The intracolumnar connections have both excitatory and feed- forward inhibitory components (Kőhler, 1985a, 1985b; Chrobak and Buzsáki, 1994; Dickson and Alonso, 1997). On the basis of the anatomical arrangement and the requirements of the mathematical reconstruction network, the model suggests that the reconstruction error (the difference between the neocortical input and the reconstructed input) is computed in layer II.

Layers V-VI of the EC are driven by the output of the CA1 field. This projection is topographically

organized (Tamamaki and Nojyo, 1990). The activities in layers V-VI of the EC are temporally integrated efferent copies (neuronal representations) of the CA1 field. Change of representation to positive coding network may occur (Lőrincz et al., 2002). Activities of layers V-VI of the EC represent the output of the hippocampus-EC system. The internal representation develops the reconstructed input for the EC by means of the layers V-VI to layer II synaptic activities. The model refers to these connections as the long-term synaptic memory of the EC. Plastic changes, underlying the synaptic memory of the EC, are brought about by a supervised tuning rule (Karhunen et al., 1997) of the CA1 outputs and the reconstruction error according to the model. The reconstruction error is developed during theta behavior, which is assumed to correspond to non-linear operation in the model. The supervisory function of the hippocampus is then operative in the linear mode, which corresponds to the sharp wave (SPW) phase.

B.5. The Output of the CA1 Region is Noise Filtered

Afferents from layer III of the EC terminate in a spatially restricted manner on the distal apical dendrites of CA1 pyramidal neurons. Through this pathway, the EC is re-mapped topographically onto the CA1 field (Tamamaki and Nojyo, 1990). It has been suggested that this pathway serves as a thresholding structure, which filters the noise of the reconstruction error (Lőrincz et al., 2002).

B.6. CA3 Recurrent Collateral Network: Replay of Learned Information During SPW

The CA3 field is active during both theta and SPW states. A major difference between the two behavioral states is that the effectiveness of the recurrent collaterals is reduced in the theta state (Hasselmo et al., 1996). According to the model, the tuning rule for the recurrent collaterals during the theta behavior is guided by the activity of the initiating CA3 pyramidal neurons and the change of firing rate of the target CA3 pyramidal neurons. In turn, the recurrent collaterals make connections between computational units of the CA3 field, they are not symmetric and may represent temporal sequences. During SPW, when the EC inputs to the CA3 region are reduced, the recurrent collateral system can 'replay' the patterns learned during the theta state. The tuning rule allows for the 'distortion' of the temporal scale during the short SPW events. As a result, the output pattern sequences of the CA3 are similar in both theta and SPW states (Wilson and McNaughton, 1994; Skaggs and McNaughton, 1996; Nádasdy et al., 1999).

B.7. The Dentate Gyrus Performs Temporal Deconvolution

Although in the model the layer V to layer II synapses are assigned as the neuronal substrate of LTM, it is also assumed in the model that the SPW-associated bursts are for tuning neocortical circuits in a fashion analogous to the training of the EC by the CA1 field (Scoville and Milner, 1957; Squire, 1992a, 1992b; McClelland et al., 1995; Chrobak and Buzsáki, 1994). Specifically, it was suggested that the outputs of the hippocampus provide some part of the training information for the neocortical hierarchy. Thus the neocortex may operate in a single, e.g., always in linear mode. Neocortical regions, however, may provide temporally convolved inputs because of improper reconstruction, for example. The temporal convolution can corrupt statistical properties and needs to be removed. The model assigned this unique role to the DG. This unique function should not be replicated in the neocortex. As the neocortex operates as a hierarchical set of linear reconstruction networks, temporal convolution that arise in neocortical networks may be removed by using component-wise deconvolution complemented with whitening and separation by the hippocampus.

Granule cells of the DG number almost one million in the rat (Claiborne et al., 1986; Seress, 1992). Thus, the internal dimension of the DG is larger than its input dimension so that the network can compress information and can form a sparse representation (McNaughton and Morris, 1987). As shown earlier (Olshausen and Field, 1996, 1997), linear response within active subspaces is possible if these subspaces are sparse. The model requires the following lossy properties of the granule cells: Losses in this network should not saturate neuronal responses but, instead, should allow approximately linear response even when the inputs are large (Olshausen and Field, 1997). In physiological terms, it predicts that the dynamic firing frequency range of granule cells is larger than that of the pyramidal neurons. It was also assumed in the model that the tuning of synapses targeting granule cell dendrites far from the somata (i.e., termination of entorhinal afferents) is suitable for whitening. Synapses targeting granule cells

close to the somata (i.e., termination of mossy cell afferents) were assumed to perform tuning limited by the lossy properties of granule cells. Non-linear saturation was assumed in the tuning but not in response properties. Moreover, as the model suggests, the response of mossy cells to granule cell excitation can be shifted in time. This assumption has been confirmed recently (Henze et al., 2001). The exact time shift may be determined by the gamma frequency population oscillation which is especially prominent in the dentate hilus (Bragin et al., 1995). Then the mossy cell afferents with approximately zero delays separate, whereas the more delayed mossy cell afferents may be able to deconvolve (Torkkola, 1996a, 1996b). The convergence-divergence pattern of the model may be described as follows. Sparse representation means that separation and deconvolution involve a small subset of granule cells, corresponding to the small number of granule cell- to mossy cell and granule cell -to CA3 pyramidal cell connections (Amaral, 1978; Acsády et al., 1998). Mossy cells can serve different granule cell subsets and thus the number of mossy cells may be small (approximately 20,000 in the rat). The divergent from mossy cell to granule cell connections (Schwartzkroin, 1994) are assigned to whit in the model.

Acknowledgments

This work was partially supported by Hungarian National Science Foundation (Grant No. OTKA 32487). Thanks are due to Dr. Edgar Körner and to the Future Technology Research Laboratory of Honda for the generous support of a parallel project on the representation of visual information. Our special thanks are due to Gyuri Buzsáki for his enlightening and continuous support during the long-course of our model construction. If it were up to us, he would be an author of this paper.

Notes

1. Here, recurrent collaterals denote within-layer associative connections to distinguish them from those in the BU flow or in the TD flow.
2. The role of matrix M is temporal integration in Fig. 2B, $\mathbf{M} = \mathbf{I}$, where $\mathbf{I}$ denotes the identity matrix.
3. This argument is based on the assumption that changes of the TD connections are negligible on the time interval of the relaxation loop.
4. For example, NMF based learning on face inputs is able to develop components that represent eyes, mouth, nose, hair, etc.
5. Novel information for one reconstruction network can be familiar for others. For example, a novel face may be processed without thresholding in the primary visual cortex, or in networks dealing with the processing of facial expressions.
6. A 'bar' is a set of ones inserted into a larger set of zeros. For visualization purposes, the ones correspond to horizontal or vertical lines if the input vector is arranged in two dimensions. See Fig. 6.
7. No layer equivalent to the IVth layer of the neocortex is present in this loop.
8. We note there can be more than one route to resolve the fallacy (see, e.g., Dennett, 1991). Along the line of the classical black box modeling the fallacy does not arise at all, but *meaningful labeling* of blocks of the model can be questioned.

References

Acsády L, Kamondi A, Sík A, Freund TF, Buzsáki G (1998) GABAergic cells are the major postsynaptic targets of mossy fibers in the rat hippocampus. J. Neurosci. 18: 3386–3403.

Adorján P, Levitt J, Lund J, Obermayer K (1999) A model for the intracortical origin of orientation preference and tuning in macaque striate cortex. Vis. Neurosci. 16: 303–318.

Amaral DG (1978) A Golgi study of cell types in the hilar region of the hippocampus in the rat. J. Comp. Neurol. 182: 851–914.

Amari S, Cichocki A, Yang H (1996) A new learning algorithm for blind signal separation. In: Advances in Neural Information Processing Systems. Morgan Kaufmann, San Mateo, CA, pp. 757–763.

Amari S, Nagaoka H (2000) Methods of Information Geometry, Vol. 191 of AMS Translations of Mathematical Monographs. American Mathematical Society, Providence, RI.

Aszalós P, Kéri Sz, Kovács Gy, Benedek Gy, Janka Z, Lőrincz A (1999) Generative network explains category formation in Alzheimer patients. In Proceedings of IJCNN 1999, Washington DC. CD ROM: JCNN2137.PDF, IEEE Catalog Number: 99CH36339C.

Bell A, Sejnowski T (1995) An information-maximization approach to blind separation and blind deconvolution. Neural Computation 7: 1129–1159.

Ben-Yishai R, Bar-Or R, Sompolinsky H (1995) Theory of orientation tuning in visual cortex. Proc. Natl. Acad. Sci. USA 92: 3844–3848.

Biederman I (1987) Recognition-by-components: A theory of human image understanding. Psychol. Rev. 94: 115–147.

Bliss TVP, Collingridge GL (1993) A synaptic model of memory: Long-term potebitiation in the hippocampus. Nature 361: 31–39.

Bousquet O, Balakrishnan K, Honavar V (1999) Is the hippocampus a Kalman filter? In: Proceedings of the Pacific Symposium on Biocomputing, pp. 619–630.

Bragin A, Jando G, Nádasdy Y, Hetke J, Wise K, Buzsáki G (1995) Gamma (40–100 Hz) oscillation in the hippocampus of the behaving rat. J. Neurosci. 15: 47–60.

Cai D, DeAngelis G, Freeman R (1997) Spatiotemporal receptive field organization in the lateral geniculate nucleus of cats and kittens. J Neurophysiol 78: 1045–1061.

Callaway E (2000) Visual cortex, cell types and connection. In: The MIT Encyclopedia of Cognitive Sciences. MIT Press, Cambridge, MA, pp. 867–869.

Calvin W (1999) Columns and modules. In: The MIT Encyclopedia of the Cognitive Sciences. MIT Press, Cambridge, MA, pp. 148–150.

Chandler D (1987) Introduction to Modern Statistical Mechanics. Oxford University Press, New York.

Charles D, Fyfe C (1998) Modelling multiple cause structure using rectification constraints. Network: Computations in Neural Systems 9: 167–182.

Chrobak J, Buzsáki G (1994) Selective activation of deep layer (V-VI) retrohippocampal cortical neurons during hippocampal sharp waves in the behaving rat. J. Neurosci. 14: 6160–6170.

Churchland P (1989) On the nature of theories: A neurocomputational perspective. In: W. Savage, ed. Scientific Theories: Minnesota Studies in the Philosophy of Science, Vol. 14 University of Minnesota Press, Minneapolis, pp. 59–101.

Claiborne BJ, Amaral DG, Cowan W (1986) A light and electron microscopic analysis of the mossy fibers of the rat dentate gyrus. J. Comp. Neurol. 246: 435–458.

Comon P (1994) Independent component analysis—A new concept? Signal Processing 36: 287–314.

Cover T, Thomas J (1991) Elements of Information Theory. John Wiley and Sons, New York, USA.

Dennett D (1987) The Intentional Stance. The MIT Press, Cambridge, MA.

Dennett D (1988) Quining qualia. In: Marcel A and Bisiach E, eds. Consciousness in Contemporary Science. Clarendon Press, Oxford, pp. 42–77.

Dennett D (1991) Consciousness Explained Little Brown, Boston, MA.

Dickson CT, Alonso A (1997) Muscarinic induction of synchronous population activity in the entorhinal cortex. J. Neurosci 17: 6729–6744.

Douglas R, Koch C, Mahowald M, Martin K, Suarez H (1995) Recurrent excitation in neocortical circuits. Science 269: 981–985.

Gluck MA (1996) Computational models of hippocampal function in memory. Hippocampus 6(6): 565–762.

Eichenbaum H (2000) A cortcal-hippocampal system for declarative memory. Nature Reviews 1: 41–50.

Field D (1987) Relations between the statistics of natural images and the response properties of cortical cells. Journal of the Optical Society of America A 4: 2379–2394.

Fodor JA (1981) Representation. The MIT Press, Cambridge, MA.

Gelder TV (1995) What might cognition be, if not computation? Journal of Philosophy XCI: 345–381.

Gluck M, Myers C (1993) Hippocampal mediation of stimulus representation: A computational theory. Hippocampus 3(4): 491–516.

Hasselmo M, McClelland J (1999) Neural models of memory. Current Option in Neurobiology 9: 184–188.

Hasselmo M, Wyble B, Wallenstein G (1996) Encoding and retrieval of episodic memories: Role of cholinergic and GABAergic modulation in the hippocampus. Hippocampus 6: 693–708.

Henze D, Wittner L, Buzsaki G (2001) Single dentate gyrus granule cells monosynaptically drive hilar and CA3 neurons in vivo. Soc. Neurosci. Abstr, vol, 27, no. 537.7.

Hinton G, Ghahramani Z (1997) Generative models for discovering sparse distributed representations. Philosophical Transactions of the Royal Society B 352: 1177–1190.

Hinton G, Sejnowski T (1983) In: Conference on Vision and Pattern Recognition, Proceedings of the IEEE Computer Society Optimal Perceptual Inference. IEEE Press, pp. 448–453.

Horn B (1977) Understanding image intensities. Artificial Intelligence 8: 201–231.

Hubel D, Wiesel T (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J. Physiol. 160: 106–154.

Hyvärinen A (1999a) Sparse code shrinkage: Denoising of nongaussian data by maximum likelihood estimation. Neural Computation 11: 1739–1768.

Hyvärinen A (1999b) Survey on independent component analysis. Neural Computing Surveys 2: 94–128.

Hyvärinen A, Hoyer P, Oja E (1999) Sparse code shrinkage: Denoising by nonlinear maximum likelihood estimation. In: Advances in Neural Information Processing Systems 11 (NIPS*98). MIT Press, Cambridge, MA, pp. 1739–1768.

Hyvärinen A, Oja E (1997) A fast fixed-point algorithm for independent component analysis. Neural Computation 9: 1483–1492.

Jensen O, Lisman J (1996) Hippocampal CA3 region predicts memory sequences: Accounting for the phase precession of place cells. Learning and Memory 3: 279–287.

Jutten C, Herault J (1991) Blind separation of sources, Part I: An adaptive algorithm based on neuromimetic architecture. Signal Processing 24: 1–10.

Káli S, Dayan P (1998) 'The formation of direction independent place fields in area CA3 of the rodent hippocampus using Hebbian plasticity in a recurrent network. Society for Neuroscience Abstracts 24: 931.

Káli S, Dayan P (1999) Spatial representations in related environments in a recurrent model of area CA3 of the rat. In: Proceedings of ICANN99.

Karhunen J, Oja E, Wang L, Vigario R, Joutsensalo J (1997) A class of neural networks for independent component analysis. IEEE Trans. On Neural Networks 8: 487–504.

Kéri S, Benedek G, Janka Z (1999) Classification learning in Alzheimer's disease. Brain (122): 1063–1068.

Kéri Sz, Janka Z, Benedek Gy., Aszalós P, Szatmaáry B, Szirtes G, Lőrincz A (2002) Categories, prototypes and memory systems in Alzheimer's disease. Trends in Cognitive Sciences 6: 132–136.

Köhler C (1985a) Intrinsic projections of the retrohippocampal region in the rat brain. I. The subicular complex. J. Comp. Neurol 236: 504–522.

Köhler C (1985b) Intrinsic projections of the retrohippocampal region in the rat brain. II. The medial entorhinal area. J. Comp. Neurol 246: 149–169.

Knowlton B, Squire L (1993) The learning of natural categories: Parallel memory systems for item memory and category-level knowledge. Science 262: 147–149.

Koch C, Poggio T (1999) Predicting the visual world: Silence is golden. Nature Neuroscience 2: 9–10.

Laheld B, Cardoso J (1994) Adaptive source separation with uniform performance. In: Signal Processing VII: Theories and Applications. Proceedings of EUSIPCO-94, Edinburgh, UK, September 1994, vol. 2. pp. 183–186.

Lee D, Seung H (1999) Learning the parts of objects by non-negative matrix factorization. Nature 401: 788–791.

Lee D, Seung H (2001) Algorithms for non-negative matrix factorization. In: Advances in Neural Processing Systems, Vol. 13. Morgan Kaufmann, San Mateo, CA.

Lee T, Bell A, Lambert R (1997) Blind separation of convolved and delayed sources. In: Advances in Neural Information Processing Systems 9 (NIPS*96). MIT Press, pp. 758–767.

Levy W (1996) A sequence predicting CA3 is a flexible associator that learns and uses context to solve hippocampal-like tasks. Hippocampus 6: 579–590.

Lisman J (1999) Relating hippocampal circuitry to function: Recall of memory sequences by reciprocal dentate-CA3 interactions. Neuron 22: 233–242.

Lisman J, Idiart M (1995) A mechanism for storing 7a2 short-term memories in oscillatory subcycles. Science 267: 1512–1514.

Lőrincz A (1997a) Common control principles of basal ganglia—thalamocortical loops and the hippocampus. Neural Network World 7: 649–677.

Lőrincz A (1997b) ICANN 97, Lecture Notes in Computer Science, Vol. 1327 In: Gerstner W, Germond A, Hasler M, Nicoud JD, eds. Hippocampal formation trains independent components via forcing input reconstruction, Springer, Berlin, pp. 163–168.

Lőrincz A (1997c) Towards a unified model of cortical computation II: From control architecture to a model of consiousness. Neural Network World 7: 137–152.

Lőrincz A (1998) Forming independent components via temporal locking of reconstruction architectures: A functional model of the hippocampus. Biological Cybernetics 79: 263–275.

Lőrincz A, Buzsáki Gy. (2000) The parahippocampal region: Implications for neurological and psychiatric diseases. In: Scharfman H, Witter M, Schwarz R, eds. Annals of the New York Academy of Sciences, Vol. 911. New York Academy of Sciences New York, pp. 83–111.

Lőrincz A, Póczos B, Szirtes G, Takács B (2002) Ockham's razor at work: Modeling of the 'homunculus'. Brain and Mind (in press).

Lőrincz A, Szatmáry B, Szirtes G, Takács B (2001a) Recognition of novelty made easy: Constraints of channel capacity on generative networks. In: French R, ed. Connectionist Models of Learning, Development and Evolution. Springer-Verlag, London, pp. 73–82.

Lőrincz A, Szirtes G, Takács B, Buzsáki Gy (2001b) Independent component analysis of temporal sequences forms place cells. Neurocomputing 38–40: 769–774.

Markram H, Lubke J, Frotscher M, Sakmann B (1997) Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. Science 275: 213–215.

McClelland J, McNaughton B, O'Reilly R (1995) Why are there complementary learning systems in the hippocampus and neocortex. Psychological Review 102: 419–457.

McNaughton BL, Morris RM (1987) Hippocampal synaptic enhancement and information storage within a distributed memory system. Trends Neurosci. 10: 408–415.

Miller K (1994) A model for the development of simple cell receptive fields and the ordered arrangement of orientation columns through activity-dependent competition between ON- and OFF-center inputs. J. Neurosci. 14: 409–441.

Miller K, Erwin E, Kayser A (1999) Is the development of orientation selectivity instructed by activity? J. Neurobiol. 41: 44–57.

Myers CE, Gluck M, Granger M (1995) Dissociation of hippocampal and entorhinal function in associative Learning: A computational approach. Psychobiology 23: 116–138.

Nádasdy Z, Hirase H, Czurkó A, Csicsvári J, Buzsáki G (1999) Replay and time compression of recurring spike sequenses in the hippocampus. Journal of Neuroscience 19: 6200–6212.

Olshausen B (1996) Learning linear, sparse factorial codes. A.I. Memo 1580, MIT AI Lab. C.B.C.L. paper no. 138.

Olshausen B, Field D (1996) Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature 381: 607–609.

Olshausen B, Field D (1997) Sparse coding with an overcomplete basis set: A strategy employed by VI? Vision Research 37: 3311–3325.

O'Reilly R, Rudy J (1999) Conjunctive representations in learning and memory: Principles of cortical and hippocampal function. Technical report, Department of Psychology, University of Colorado, Boulder.

Oussar Y, Dreyfus G (2001) How to be a gray box: Dynamic semi-physical modeling. Neural Networks 14: 1161–1172.

Paatero P, Tapper U (1994) Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. Environmetrics 5: 111–126.

Rao R, Ballard D (1999) Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. Nature Neuroscience 2: 79–87.

Riesenhuber M, Poggio T (1999) Hierarchical models of object recognition in cortex. Nat Neurosci. 2: 1019–1025.

Rolls E (1989) Functions of neuronal networks in the hippocampus and neocortex in memory. In: Byrne JH, Berry WO, eds. Neural Models of Plasticity: Experimental and Theoretical Approaches. Academic Press, San Diego, pp. 240–265.

Rolls E (2000) Hippocampo-cortical and cortico-cortical backprojections. Hippocampus 10: 380–388.

Saul A, Humphrey A (1990) Spatial and temporal response properties of lagged and nonlagged cells in cat lateral geniculate nucleus. J Neurophysiol. 64: 206–224.

Schwartzkroin PA (1994) Role of the hippocampus in epilepsy. Hippocampus 3: 239–242.

Scoville W, Milner B (1957) Loss of recent memory after bilateral hippocampal lesions. J. Neurol. Neurosurg. Psychiatry 20: 11–21.

Searle J (1992) The Rediscovery of Mind. Bradford Books, MIT Press, Cambridge, MA.

Seress L (1992) Morphological variability and developmental aspects of monkey and human granule cells: Differences between the rodent and primate dentate gyrus. Epilepsy Res. Suppl. 7: 3–28.

Simoncelli E, Olshausen B (2001) Natural image statistics and neural representation. Annual Review of Neuroscience 24: 1193–1215.

Skaggs WE, McNaughton BL (1996) Replay of neuronal firing sequences in rathippocampus during sleep following spatial experience. Science 271: 1870–1873.

Smolensky P (1989) Connectionism and the language of thought. In: Loewer B, Rey G, eds. Meaning in Mind: Fodor and His Critics. Basil Blackwell Ltd. Oxford, pp. 201–227.

Somers D, Nelson S, Sur M (1995) An emergent model of orientation selectivity in cat visual cortical simple cells. J. Neurosci. 15: 5448–5465.

Squire L (1992a) Declarative and nondeclarative memory: Multiple brain systems supporting learning and memory. J. Cog. Neurosci. 4: 232–243.

Squire L (1992b) Memory and the hippocampus: A synthesis of findings with rats, monkeys, and humans. Psychol. Rev. 99: 195–231.

Steinvas I, Intrator N, Moshaiov A (2000) Improving Classification via Reconstruction. Available at citeseer.nj.nec.com/article/steinvas00improving.html.

Suarez H, Koch C, Douglas R (1995) Modeling direction selectivity of simple cells in striate visual cortex within the framework of the canonical microcircuit. J. Neurosci. 15: 6700–6719.

Szirtes G, Lőrincz A (2002) Low level priming as a consequence of perception. In: Builinaria A, Lave W, eds. Proceedings of the Seventh Neural Computation and Psychology Workshop (NCPW7). World Scientific, pp. 223–235.

Tamamaki N, Nojyo Y (1990) Disposition of the slab-like modules formed by axon branches originating from single CA1 pyramidal neurons in the rat hippocampus. J. Comp. Neurol. 291: 509–519.

Torkkola K (1996a) Blind separation of convolved sources based on information maximization. In: IEEE Workshop on Neural Networks for Signal Processing Acoustics, Speech and Signal Processing, Kyoto, Japan. pp. 423–432.

Torkkola K (1996b) Blind separation of delayed sources based on information maximization. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing. pp. 3510–3513.

Troyer T, Krukowski A, Priebe NJ, Miller KD (1998) Contrast-invariant orientation tuning in cat visual cortex: Feedforward tuning and correlation-based intracortical connectivity. J. Neurosci. 18: 5908–5927.

Wilson M, McNaughton B (1994) Reactivation of hippocampal ensemble memories during sleep. Science 265: 676–679.

Wimbauer S, Wenish O, Miller K, Hemmen van J (1997) Development of spatiotemporal receptive fields of simple cells: I. Model formulation. Biological Cybernetics 77: 456–461.