Neural Networks 22 (2009) 738–747

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

2009 Special Issue

Here and now: How time segments may become events in the hippocampus

András Lőrincz *, Gábor Szirtes

Department of Software Technology and Methodology, Eötvös Loránd University, Pázmány sétány 1/C, Budapest, H-1117, Hungary

ARTICLE INFO

Article history:
Received 6 May 2009
Received in revised form 25 May 2009
Accepted 25 June 2009

Keywords:
Place cells
Grid cells
Hippocampus
Independent Process Analysis
Time series model

ABSTRACT

The hippocampal formation is believed to play a central role in memory functions related to the representation of events. Events are usually considered as temporally bounded processes, in contrast to the continuous nature of sensory signal flow they originate from. Events are then organized and stored according to behavioral relevance and are used to facilitate prediction of similar events. In this paper we are interested in the kind of representation of sensory signals that allows for detecting and/or predicting events. Based on new results on the identification problem of linear hidden processes, we propose a connectionist network with biologically sound parameter tuning that can represent causal relationships and define events. Interestingly, the wiring diagram of our architecture not only resembles the gross anatomy of the hippocampal formation (including the entorhinal cortex), but it also features similar spatial distribution functions of activity (localized and periodic, 'grid-like' patterns) as found in the different parts of the hippocampal formation. We shortly discuss how our model corresponds to different theories on the role of the hippocampal formation in forming episodic memories or supporting spatial navigation. We speculate that our approach may constitute a step toward a unified theory about the functional role of the hippocampus and the structure of memory representations.

© 2009 Elsevier Ltd. All rights reserved.

1. Introduction

Although our senses receive an enormous amount of information at every time instant, we have the remarkable ability to filter out, organize and store only those pieces of information that might be relevant from behavioral, physical or cognitive aspects. In addition, sensory information processing is believed to facilitate *prediction* (Bialek, Nemenman, & Tishby, 2001) of behaviorally relevant changes of observations including both internal and external variables. How this prediction actually works is an open question, but it is generally assumed that it is based on the creation of internal representations of lower complexity. In the temporal domain such condensed representation may lead to the notion of events. An *event* may intuitively be defined as a primary cause that results in temporally bounded change of a given state or condition behind the observations. For example, an animal may be still or moving fast, and anything (detecting a predator or a potential mating partner) that can trigger a switch between these states could be seen as an event. By learning causal relationships between events and the resulting changes, it becomes possible to predict succeeding states by detecting a particular event. However, in contrast to the widely used coding mechanisms where codewords can easily be distinguished, we receive a continuous flow of sensory signals. Due to the limited memory capacity, an efficient segmentation mechanism is required to help encode the incoming signals. In Section 2 we formalize our assumptions on the sensory signals and – based on the notion of *statistical independence* – we show how the resulting model can be used for segmentation. Statistical independence is often coupled with sparse coding (Földiák, 2002; Olshausen & Field, 1997; Seeger, 2008), which has been suggested as the underlying neural mechanism for optimal reconstruction regarding redundancy reduction and stimulus reconstruction. What it means is that most (natural) stimuli can be decomposed into a finite set of features which are sparse (that is they are 'silent' in most of the time), but when they can be detected, their contribution to the overall signal is quite important. This sparsity can then be used as a time stamp that marks the start and end points of state transition processes. In Section 3 we translate the algorithms into a connectionist network in which parameter tuning can only be realized via biologically plausible local interactions. In Section 4 we briefly sketch the functional parallels between our computational model and the hippocampal region (HR). For decades, hippocampus research has been dedicated to study either its role in episodic memory or its role in spatial navigation (Eichenbaum, 2000; O'Keefe & Nadel, 1978; Vargha-Khadem et al., 1997). Recently, a new line of research has emerged that attempts to unite the seemingly orthogonal theories developed in these two parallel tracks (e.g. Mizumori (2006)). Our model is essentially a signal encoding system dealing with time series and representations; thus it seems to belong to the first venue. To see how it behaves when there is space related

* Corresponding author. Tel.: +36 20 324 2453; fax: +36 1 381 2140.
E-mail addresses: andras.lorincz@elte.hu (A. Lőrincz), szirtes@inf.elte.hu (G. Szirtes).

0893-6080/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neunet.2009.06.020

information in the sensory signals (which is of fundamental importance in spatial navigation problems), we present some simulation results in Section 5 about the dynamics of the model when applied on inputs with explicit spatial dependence. Finally, in Section 6 we shortly discuss the results as well as the limitations of the current model and propose a mechanism that may support and extend our model even if such explicit spatial dependence of the inputs cannot be assumed. One of the most interesting issues is the relation between our information theoretically motivated model and theories on the sensory-motor integration process proposed to form predictive internal models of the world. We also formalize some predictions on the functioning and dynamics of the HR. A short version of this paper with a different emphasis has been accepted at IJCNN'09 (Lőrincz & Szirtes, 2009).

2. Identification of the sensory input

It is natural to interpret the observations ($x(t) \in \mathbf{R}^d$) as mixed signals emitted by a hidden (not directly observable) state variable ($s(t) \in \mathbf{R}^d$) that evolves in time thus forming a process. The simplest case is if linearity is assumed both for the mixing and the dynamics of the process. Accordingly, the observations and the hidden process may be written as:

$$x(t) = As(t)$$
$$s(t+1) = \sum_{i=0}^{I-1} F_i s(t-i) + \sum_{j=0}^{J-1} H_j e(t+1-j)$$

that is the observations are instantaneous mixtures of the state components whose evolution follows an autoregressive moving average process (ARMA) of order (I, J) with driving noise (or innovation process) $e(t) \in \mathbf{R}^d$. In general the driving noise components are assumed to be *temporally* independent and identically distributed (i.i.d.) stochastic variables. However, in accord with our causal definition of events we also assume that noise components (or at least their subgroups) are *spatially* (i.e., index-wise) independent. We assume that matrix $H_0 \in \mathbf{R}^{d\times d}$ is the identity matrix $I_d \in \mathbf{R}^{d\times d}$. The ARMA$(I, J)$ model comprises the contributions of previous states transferred by the predictive matrices F_i $(i = 0, \ldots, I-1)$ and the different echoes of the driving noise transferred by matrices H_j $(j = 0, \ldots, J-1)$. The goal is to find the $\hat{F}_i \in \mathbf{R}^{d\times d}$ $(i = 0, \ldots, I-1)$ and $\hat{H}_j \in \mathbf{R}^{d\times d}$ $(j = 0, \ldots, J-1)$ estimations of the hidden dynamics and the echo structure, respectively, and to learn to separately represent the estimated hidden state, $\hat{s}(t)$, and the independent subgroups of the estimated driving noise $\hat{e}(t)$.

Regarding the structure of F_i, $i = (0, \ldots, I-1)$, a special case seems relevant: *joint block-diagonal structure* implies dynamical sub-processes that do not *mix*. These hidden processes are independent in a dynamical sense so their identification could reduce the representational complexity.

For simplicity, we assume a hidden ARMA$(1, 0)$ = AR(1) process and the issue of delays will be discussed later. (On higher order hidden ARMA processes and post-nonlinear extensions see Lőrincz and Szabó (2007), Szabó, Póczos, and Lőrincz (2007a, 2007b) and Szabó, Póczos, Szirtes, and Lőrincz (2007), respectively.)

The key step to solve the identification problem is to recognize that the observation process is also an AR(1) process, if matrix A can be inverted:

$$x(t+1) = Mx(t) + n(t+1), \tag{1}$$

where the observation noise is $n(t+1) = Ae(t+1)$. According to the d-central limit theorem (Petrov, 1958) $n(t+1)$ is approximately Gaussian so the predictive matrix $(M = AFA^{-1})$ may be estimated by least-mean square approximations and then the wanted independent driving noise components can be extracted by applying Independent Component/Subspace Analysis (ICA/ISA) (Cardoso, 1998; Comon, 1994; Jutten & Herault, 1991) on the observation noise. An important result (Póczos, Szabó, Kiszlinger, & Lőrincz, 2007; Póczos & Lőrincz, 2006) is that – for a large class of source distributions – separation of independent subspaces (i.e., the ISA problem) can be solved in two steps. First, traditional ICA methods yield one dimensional components and second, the resulting components should be grouped to form independent subspaces. In turn, ISA of the estimated noise can simultaneously recover the estimated mixing matrix $\hat{A}$, the hidden state $\hat{s}(t) = \hat{A}^{-1}x(t)$ as well as the assumed independent subspaces of the multidimensional driving source components $\hat{e}(t) = \hat{A}^{-1}\hat{n}(t)$. If $\hat{A}$ is recovered, then the hidden predictive matrix can also be approximated ($\hat{F} = \hat{A}^{-1}\hat{M}\hat{A}$).

3. A connectionist network implementation

In this section we provide local learning rules for parameter estimation of the identification task described above. The resulting algorithms can be translated into a neural network in which 'activity' of a 'neuronal layer' is represented by a vector, connection weights between layers or within layer components (i.e., recurrent connections) are represented by matrices and neurons may realize nonlinear transformations of their inputs.

3.1. Assumptions and simplifications

For simplicity, rate coding (manifesting analogue values) and mixed, i.e., positive and negative weights (thus contradicting with Dale's Principle) are assumed throughout the derivations, but the proposed functioning can in principle be also realized by using either positive coding (Plumbley, 2002) or homogeneous connection systems (Parisien, Anderson, & Eliasmith, 2008). Inhibitions (or subtractions) are manifested by separate inhibitory populations within a layer using feedback or feed-forward inhibition.

We also assume that after each transformation the resulting entities (represented by a given layer) get decorrelated and normalized (i.e. they are subject to *whitening*). Whitening helps compressed encoding (Brand, 2006) and it speeds up ICA (Amari, Cichocki, & Yang, 1996; Cardoso & Laheld, 1996) if applied during preprocessing. Whitening may take place within a layer with the help of inhibitory recurrent connections for which biologically feasible learning rules can also be given (Földiák, 1989). Interestingly, whitening can also be realized by utilizing inter-layer feedback connections. The so called forward-inverse model (Kawato, Hayakawa, & Inui, 1993) or reconstruction network (Lőrincz & Buzsáki, 2000) assumes a *loopy* structure with a *bottom-up* (BU) and a *top-down* (TD) transformations in which representation of the input is used to regenerate an estimate of the input and different constraints can be put on the transformations. It can be shown that in a reconstruction network the two branches can learn to invert each other – at least in the pseudo-inverse sense – applying simple Hebbian rules that we introduce later. Both whitening and ICA can be realized in a reconstruction network for which the mathematical details and a possible implementation scheme will soon be provided.

The different algorithmic tasks and the related learning rules will be described in separate subsections below.

3.2. Innovation

The first algorithmic step is the estimation of the innovation: $\hat{n}(t+1) = x(t+1) - \hat{M}x(t)$. The corresponding cost function $J(\hat{M}) = \frac{1}{2}\sum_t |x(t+1) - \hat{M}\,x(t)|^2$ leads to the following Hebbian

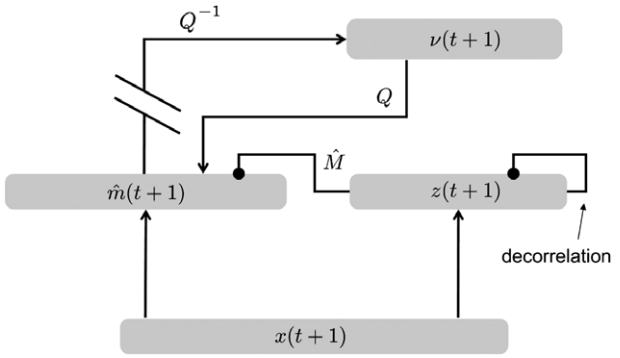

Fig. 1. Connectionist architecture that learns AR dynamics by Hebbian learning and returns whitened innovation.

learning rule: $\Delta\hat{M}(t+1) \propto \hat{n}(t+1)x(t)'$, where $'$ denotes transpose. Since this rule requires simultaneous access to the error and the observation, the realization of this rule assumes two separate layers representing the observation and the innovation, respectively. As we see, while the learning rule is based on positive correlations, but the transformation itself is a subtraction, so $\hat{M}$ is supposed to act through feed-forward inhibition. It also implies that observation is projected onto both layers. To fulfill the whitening requirement on the represented entities, both the architecture and the learning rule need to be modified. Let $z(t)$ and $\hat{m}(t)$ denote the whitened observation and the whitened estimated innovation, respectively. Distinct whitening procedures are needed for the two signals. Observation can be made white by intra-layer inhibitory collaterals. Innovation, however, may *also* get whitened by integrating the two layers into a loop structure. The effect of the loop is represented by an auxiliary term in the cost function:

$$J(\hat{M}) = \frac{1}{2}\sum_t |x(t+1) - \hat{M}\,z(t) + Q\,\nu(t+1)|^2. \tag{2}$$

The exact role of $Q \in \mathbf{R}^{d\times d}$ and $\nu(t+1) \in \mathbf{R}^d$ will be given later. For now it is enough to conceive $\nu(t+1)$ as an internal representation of the innovation created by a reconstruction network (see, Fig. 1). The advantage of the auxiliary connection system Q is that learning rule for Q

$$\Delta Q(t+1) \propto [Q(t)']^{-1} - \hat{m}(t+1)\nu(t+1)' \tag{3}$$

can make the activity at the innovation layer white (Cardoso & Laheld, 1996). This learning rule can be strictly Hebbian by means of *effective coincidences* (Lőrincz & Buzsáki, 2000) (see later). In addition, the online learning rule of $\hat{M}$ still assumes a pure Hebbian form:

$$\Delta\hat{M}(t+1) \propto \hat{m}(t+1)z'(t). \tag{4}$$

3.3. Separation

The next module extracts the independent noise components from the innovation: $\hat{e}(t) = W\hat{m}(t)$. Among several solutions that can estimate *demixing matrix* W, the so called Infomax online rule (Bell & Sejnowski, 1995) suits our architecture:

$$\Delta W(t+1) \propto \left[W(t)'\right]^{-1} - f\left(\hat{e}(t)\right)\hat{m}(t)'. \tag{5}$$

Here, $f(.)$ denotes tangent hyperbolic component-wise nonlinearity. This rule is actually a nonlinear version of Eq. (3) and locality can also be ensured by the loopy structure. Let us remark that pre-whitening significantly accelerates Eq. (5) (Amari et al., 1996; Cardoso & Laheld, 1996). In order to couple the hidden states with the corresponding driving noise, a transformation that matches indices and compensates for differences of pre-processing steps should be learnt on the white observation: $\hat{s}(t) = \tilde{W}z(t)$. The trivial solution is if $\hat{m}$ and an estimation of z are channeled through the same connection system. Another solution is if the learnt transformation is 'mirrored' onto a parallel BU channel thus the two signals are transmitted separately. As we argued in Lőrincz and Szirtes (2009) both solutions may be realized in a neurally plausible way at the same time, offering a robust solution to this delicate problem.

3.4. Learning to predict

The *excitatory* predictive matrix of the hidden process ($\hat{F}$) can be learnt with the help of the recovered independent driving noise components and the estimated hidden states. To comply with the whitening requirement, the internal predictive model is supposed to work on white signals. The resulting model could be used to predict future hidden states. We propose a separate layer with a recurrent network that receives two inputs; the estimated driving noise ($\hat{e}$) and the estimated hidden state ($\hat{s}$) channeled through the same connection system, but in a temporally multiplexed *two-phase* manner. Let h denote the (intra-layer) whitened hidden states. The prediction of future hidden states assumes the following dynamics:

$$h(t+1) = \hat{F}h(t) + \nu(t+1). \tag{6}$$

Transformed driving source $\nu(t+1)$ acts as an error signal, so we get a supervised Hebbian learning rule:

$$\Delta\hat{F}(t+1) \propto \nu(t+1)h(t)' \tag{7}$$

which minimizes the cost function

$$J(\hat{F}) = \frac{1}{2}\sum_t |h(t+1) - \hat{F}\,h(t)|^2. \tag{8}$$

Although computationally such rule is feasible, the neural implementation is not straightforward. A plausible mechanism is given in the next section.

4. Functional mapping

We argue that our proposed learning model may describe an important aspect of episodic memory learning in the HR. As it was detailed in Lőrincz and Szirtes (2009), our claim is based on a set of features unique to the neural substrate. Here we only present the functional correspondences between the two systems. To help the comparison, the gross anatomy of the mammalian hippocampal region (comprising the entorhinal cortex (EC), subfields CA3 and CA1, dentate gyrus (DG), para and presubiculum and the subiculum (Witter & Amaral, 2004)) and the wiring diagram of our connectionist architecture are depicted in Fig. 2(a) and (b), respectively.

The superficial entorhinal layers are supposed to receive inputs (highly processed sensory stimuli and motion related activity) from the cortex. Due to the proximity of EC layers II and III, they may share the same input thus providing the parallel pathways required by the innovation step. As EC II has an extensive inhibitory network, we speculate that connections from EC III mainly act through feed-forward inhibition and EC II represents innovation. The recurrent connections in both layers may take part in whitening, because the diversity of inhibitory populations may enable concurrent functions without interference. In addition, projections from EC deep layers may also contribute to whitening of EC II activity. In Lőrincz and Buzsáki (2000) the loops within tri-synaptic pathway have been assigned to handle the MA contributions caused by the delays in the architecture. For simplicity, now we only discuss AR processes.

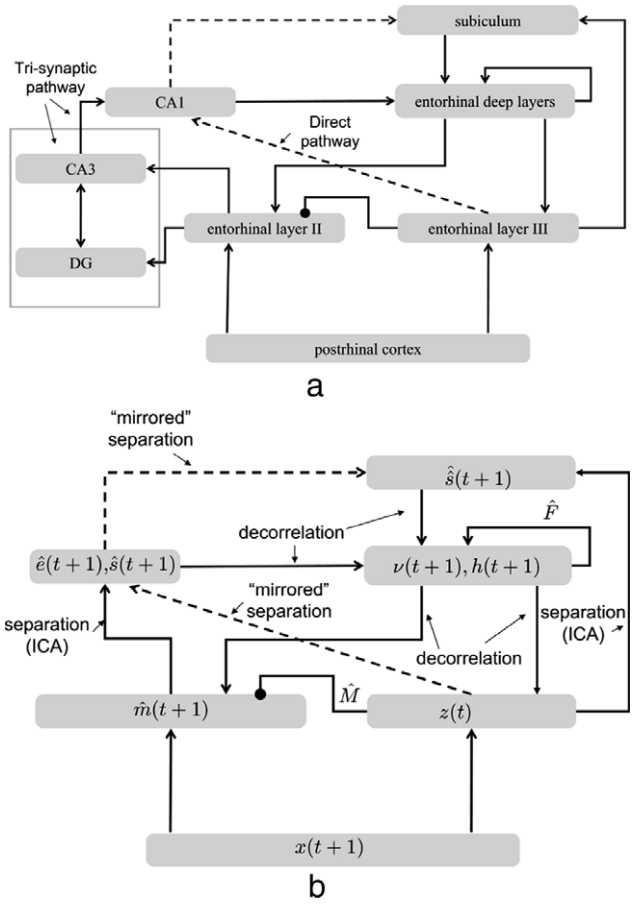

Fig. 2. (a) The main wiring diagram of the hippocampal region. For the sake of completeness, loops between the dentate gyrus (DG) and CA3 are also depicted. In our model, the putative role of these loops is to remove large delays (moving averages). (b) Connectionist architecture for representing the independent noise and state recovered from the observation signals. Dashed lines denote connections tuned in a supervised manner. Arrowhead lines denote mostly excitatory, circle head denotes inhibitory connections. Note the two ICAs giving rise to $\hat{s}$ and $\hat{\hat{s}}$, respectively.

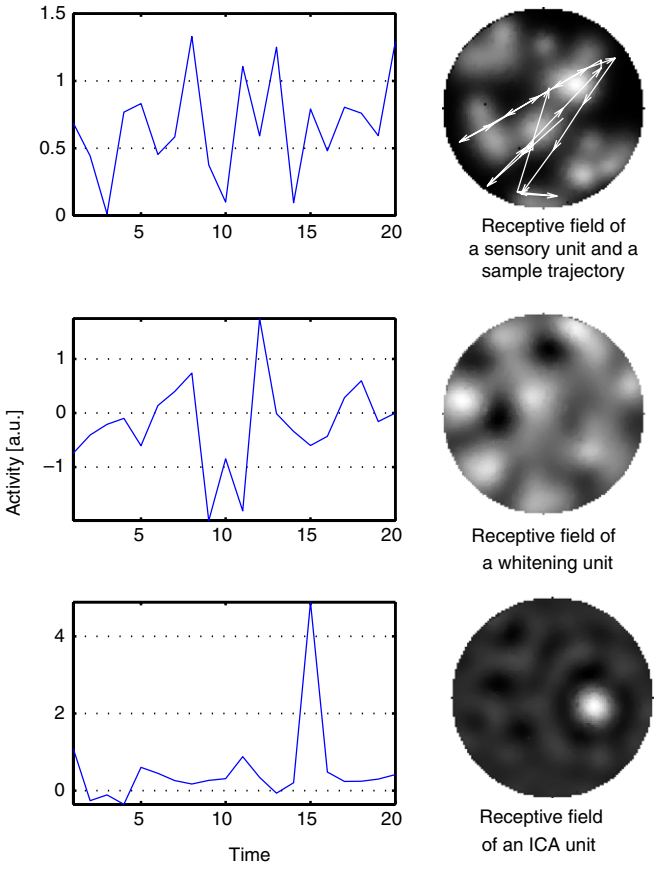

Fig. 3. Sample time series of sensory, whitening and separating units. Left column: time series of activity of units of different layers. Right column: the corresponding receptive fields (without normalization and thresholding). Top row: activity (in arbitrary units) of a sensory unit. The receptive field features multiple peaks. The corresponding sample trajectory is superimposed on the receptive field. Middle row: activity (in arbitrary units) of a whitening unit corresponding to the sample trajectory. The receptive field is more regular ('grid cell'). Bottom row: activity (in arbitrary units) of a unit of the separation layer corresponding to the sample trajectory. The receptive field is localized ('place cell'). The diameter of the maze is 2m, the size of the 'virtual rat' is about 55 cm and the sampling distance along the trajectory is also 55 cm.

Accordingly, connections between EC II and CA3 through DG (the whole tri-synaptic pathway) are collapsed and treated as a single synaptic channel (Fig. 2(b)). In turn, the superficial layers communicate with CA1 through two parallel channels that transmit the estimated independent noise (separation process) and the transformed states (*mirrored* separation), respectively.

The supervised training of the hidden model requires the matching between the noise and state components. In turn, separation matrices between EC II to CA1 and EC III to CA1 should be tightly coupled in spite of the potential differences of the whitening transformations acting on EC II and EC III and the permutation invariance of ICA. We speculate that this *index matching* of the two learning processes can be realized by 'cooperative learning' between the distal and proximal synapses of CA1 (Golding, Staff, & Spruston, 2002; Mehta, Natarajan, Tadepalli, & Fern, 2005). We also speculate that CA3 is able to transmit different signals in different phases of theta. When its recurrent connections are suppressed, it transmits the innovation and tuning may be governed by Eq. (5) yielding the separation transformation. However, when the collaterals are active, CA3 can integrate the previous and current innovations to signal an estimate of the observation. In this phase, CA3-to-CA1 transformation yields the estimated hidden state which serves as a *supervisory* output of the *other* channel thus substituting the role of the post-synaptic activity in Hebbian learning. Nevertheless the direct route shares common properties with the tri-synaptic route and may also learn to separate via learning rule (5) unless supervisory CA3 signals override it. This proposal is in accord with the findings (Leutgeb, Leutgeb, Treves, Moser, & Moser, 2004) that approximate place fields emerge first in CA1, but place fields of CA3 cells get stabilized faster. When the parallel channels are correctly tuned, then their transformed inputs are adjusted to each other. We also conjecture that in one phase of the theta wave CA1 transmits the estimated hidden driving sources and in the other phase it transmits the estimated hidden states. Since the components of the latter term may not be independent it implies that CA1 output collected during the two phases should show different statistics.

Another critical issue about the separation is that the necessary synaptic enhancement of separation (regarding the term $[W']^{-1}$ in Eq. (5)) cannot be realized by a recurrent connection system as it is absent in CA1. Our proposal is built on the reconstruction network idea utilizing loopy connections. As Eq. (5) shows, the second Hebbian term directly depends on the pre- and post-synaptic activities. However, the first term is proportional to the inverse transpose of the transformation matrix ('top-down' (TD) connections) which basically equals the rest of the loop. Since the reconstruction network is aiming to regenerate the input, the reconstruction error decreases as parameter tuning advances. In

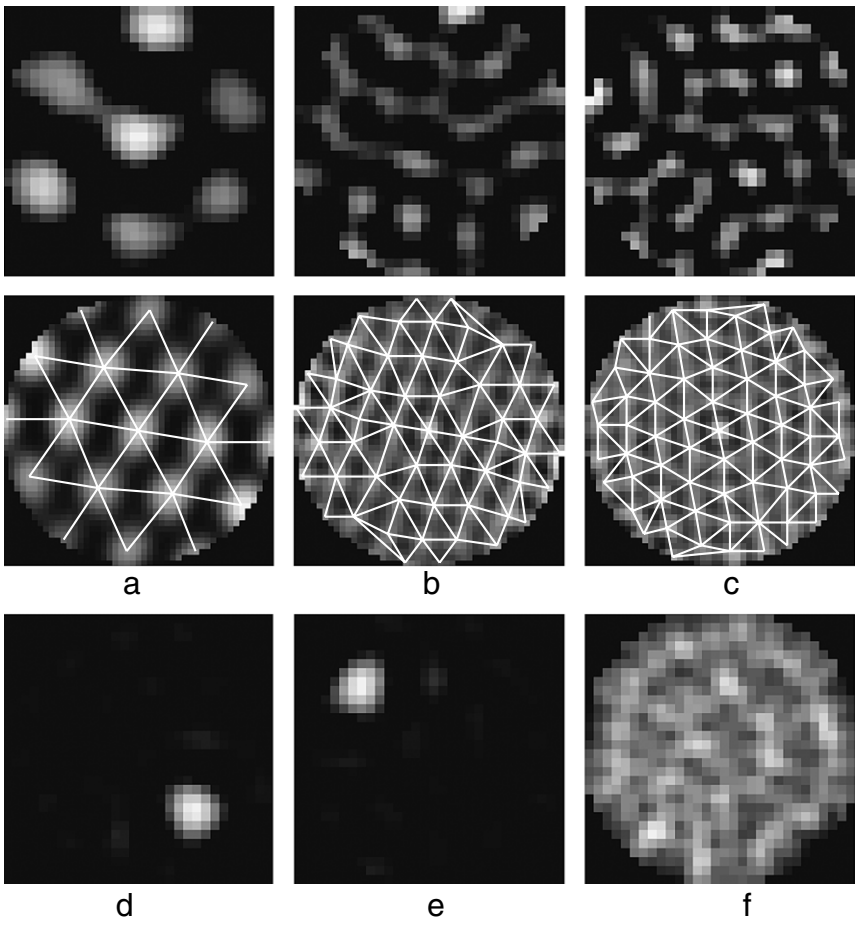

Fig. 4. Position dependent input. (a)–(c): each column shows the output of different *decorrelating* units (whitening). First row: half-wave rectified and scaled activity maps (0: black, 1: white). Second row: two dimensional autocorrelation function of the activity maps and the fitted grids. (d) and (e): sign corrected activity maps of two *separating* (ICA) units. Response is localized. (f): superimposed map of all ICA units demonstrating that the localized units cover the full maze.

turn, this TD term will transmit mostly noise for which the so called *effective coincidences* can maintain locality. Such coincidences may occur within the time window while Ca^{2+}-permeable N-methyl-D-aspartate channels or voltage-gated Ca^{2+} channels are open (Lőrincz & Buzsáki, 2000). This proposal is in accord with the anatomy of HR, because CA1 projects onto the deep layers of EC, which in turn project onto the input (superficial) layers of EC.

EC deep layer to superficial layer transformations may also take part in the whitening of the activity at the superficial layers. In turn, the auxiliary term $Q\nu(t+1)$ of Eq. (2) may augment intra-layer whitening through recurrent collaterals at the superficial layers, provided that noise of the model network is transmitted top-down in one of the theta phases.

The subiculum as one of the chief projection areas of CA1 is not modeled now, although we hypothesize that the special cross connections found in CA1 and subiculum (Gigg, 2006) are used to create a second separation loop in order to have a distinct place for representing hidden states $\hat{s}$ denoted by $\hat{\hat{s}}$ in Fig. 2(b), because – in the absence of another index matching process – they may differ. Information from CA1 and subiculum are optimized for grouping them into independent subspaces for the hidden causes and into independent deterministic processes for the hidden states. Such grouping may occur at the deep layers of EC as well as at other parts of the brain including the frontal lobe.

5. Simulation results

In our simulations the 'rat' 'runs' in a 2m wide, open-field circular arena on a linear path at a constant speed and may make a random turn at each step. It also makes a random turn when possible collision with the wall is detected. Input sampling has been fixed to 55 cm. A sample trajectory is shown superimposed on the receptive field (activity map) of a typical sensory unit on the top row in Fig. 3.

The most restrictive approximation in our simulations is that the inputs to the HR do not contain information about distant cues. This simplification enabled us to avoid the arbitrariness in modeling low-level sensory processing, because we could simply mimic postrhinal (Burwell & Hafeman, 2003) inputs. It also implies that distortions of spatial activity distributions as a function of maze distortions (Barry, Hayman, Burgess, & Jeffery, 2007) becomes trivial in our case (see Section 6).

Sensory information has been sampled along the trajectory and then analyzed by Auto-Regressive Independent Process Analysis methods (Póczos et al., 2007; Póczos & Lőrincz, 2006). The following three types of inputs have been used: (1) 1000 mixed local cues (similar to the one shown in the right hand side of top row of Fig. 3), (2) mixed local cues with units that have directional sensitivity without locality and (3) conjunctive inputs, where units show spatial *and* directional selectivity.

Fig. 3 displays different time series of activity corresponding to sensory, whitening and the separating units (without normalization and thresholding). It can be seen that – in contrast to the

activity of the whitening units – the extracted independent components are sparse indeed, so their presence can mark the start and end points of state transitions behind the continuous observations. In turn, according to our working hypothesis, these sparse signals can be used to denote the wanted events and to segment the flow of information.

As opposed to real spiking data, linear transformations may give rise to negative signals. In turn, the correspondence between the unit activity values after each transformation and the neurons' response is not straightforward. For generating the activity maps of the input units, first we discretized the space (the resolution was 30 × 30 so a bin is 6.67 cm × 6.67 cm, which is comparable of the experimental data (Hafting, Fyhn, Molden, Moser, & Moser, 2005a), and for each bin we summed up the activity measured in those steps that ended in the given bin. This spatial averaging smoothes out the artifacts caused by unattended spots. The activity after, e.g., decorrelation may assume negative values so data were rectified and scaled to range [0, 1]. Since separation is invariant for the change of sign (Jutten & Herault, 1991), the following procedure has been applied: we ranked the bins by their absolute value, and the the sign of the majority of first 10 bins defined the sign of the activity map. Then as it was done before, negative values have been clipped and the rectified data have been scaled onto range [0, 1]. We also computed the two dimensional normalized autocorrelation for each activity map. For spatial analysis of the peak activity regions of the autocorrelation image, we fitted a grid on the locally maximal points using Delaunay-triangulation (Markus et al., 1995; Takács & Lőrincz, 2007).

For the case of local cues, whitening yielded reasonable hexagonal grid structures (Fig. 4(a)–(c)). ICA yielded units with 'place cell' like, localized activity (Fig. 4(d), (e)). The size of the cells depended on the number of ICA components; nevertheless they always covered the arena uniformly (Fig. 4(f)).

The increased regularity in comparison with the original inputs can be seen by the peaky distribution of the mean edge lengths (Fig. 5(a)). The distribution of the hexagrid angles also shows the increased regularity. Angles between the vertices were close to 60 degrees (Fig. 4(a)–(c)) and their standard deviation was less than 20 degrees, thus yielding a much narrower distribution than that of the random grids (Fig. 5(b)).

When the input included direction dependent and space dependent 'sensors', whitening resulted mostly in direction dependent, but still grid-like spatial activity (Fig. 6(a–c)). In a few cases, decorrelation yielded units with clear direction dependence without any regular spatial dependence (not shown). On the other hand, ICA yielded two well separable groups of units. In one set, units show clear spatial dependence, i.e., rotationally invariant, place cell like activity (e.g. Fig. 6(d)). In the other set, the units show global direction dependence without any obvious localization; similar to head direction cells (Fig. 6(e–f)).

For the case of conjunctive inputs, whitening and ICA were both direction dependent (we show the results for ICA in Fig. 7). We computed the innovation along the straight lines of the trajectories that decreased the directional dependence considerably. (Compare the two sets of polar plots in Fig. 7.).

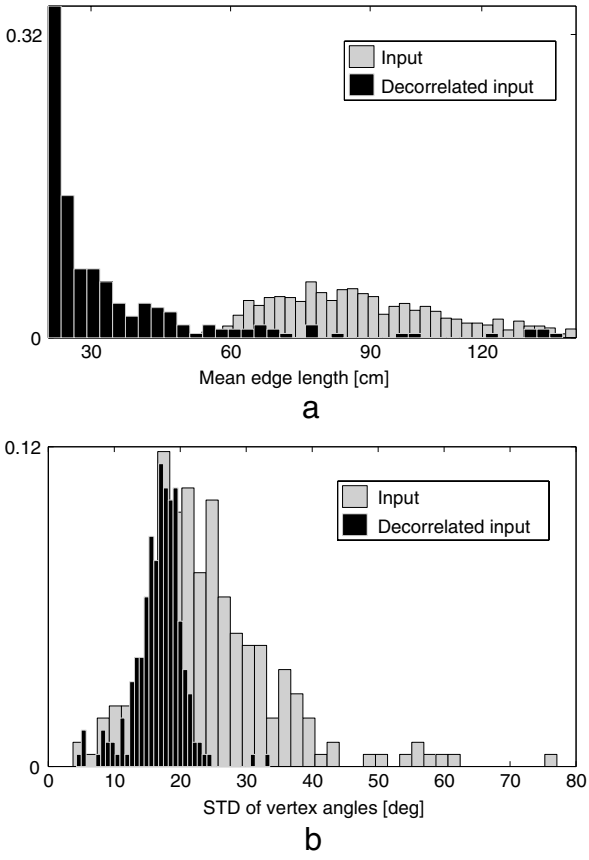

Fig. 5. Position dependent input. (a): histogram of the mean edge length of the grids for the input set and the whitening units. (b): histogram of the standard deviation of the vertex angles for the input set and the whitening units.

6. Discussion

In this paper we argue that events stored in episodic memory should refer to spatially and temporally bounded changes in the hidden state behind the observations about the environment. Events have an intrinsically finite structure in contrast to the continuous sensory signals they are derived from. Our solution to this issue is based on modeling the sensory signals as a hidden ARMA process driven by independent sources. For continuous signals, independence usually is coupled with sparse activity, so the sources can naturally act as time stamps for events. In addition, extraction of *multidimensional* independent sources requires extraction and grouping of the one dimensional components (cf. ISA). We hypothesize that *CA1's unique function* is to recover the simple sources which then are grouped (second step of Independent Subspace Analysis, (Póczos et al., 2007; Szabó et al., 2007a)) and integrated by EC as well as the neocortex. This grouping is then an integral part of memory consolidation (on systems consolidation, see Nadel, Winocur, Ryan, and Moscovitch (2007)).

Description of events, however, is not complete without defining a corresponding time scale. Changes at a given time scale may trigger other changes at different time scales, thus defining a *hierarchy* of events. In our simulations signals interpreted as triggers have shown place cell like activity. Due to the coarse sampling and the lack of distal cues, place fields are always traversed in about one time step (the size of place fields depends on the number of ICA components). In real experiments, however, place cells are active for several theta periods. In turn, our model corresponds to dynamics with slow triggers present for a time span compatible with the animal's speed instead of the sampling rate.

We have also translated the proposed algorithmic model into a connectionist framework in which biologically motivated (Hebbian or local) interactions are allowed for parameter learning. As it was detailed in Lőrincz and Szirtes (2009) a functional mapping can be given between the resulting architecture and the hippocampal region. Although our model has been derived from first principles (like redundancy reduction and Independent Process Analysis) and constrained by functional and computational consideration (as described in Section 3), it may explain several

Fig. 6. Mixture of clearly position and clearly direction dependent inputs. (a–c): each column shows the output of different whitening units. First row: two dimensional autocorrelation function of the activity maps and the fitted grids. Second row: overall direction selectivity depicted on a polar plot. (d–f): each column corresponds to 3 sample output units of the *separation*. First row: half-wave rectified and scaled activity maps as in Fig. 4. Second row: overall direction selectivity depicted on a polar plot.

peculiarities of the HR, often overlooked by other models (e.g. the requirement of parallel loops between the entorhinal cortex and CA1, two phase working of HR and the different inhibitory connection systems of EC II and EC III). When our signal encoding model is applied on spatially defined inputs, the 'emergent' (i.e., not designed in our ARMA-IPA description) spatial behavior of the different modules resembles the ones featured by the corresponding parts of the HR thus giving additional support to our functional mapping. This resemblance may give rise to a different interpretation of the place cells. They may not describe space per se, but they represent a set of observations that trigger given changes of the hidden state (i.e., they drive the hidden process) described by the learnt internal model. Also, events may receive a spatial label from the outputs of these cells.

The emergence of these spatial activity patterns is relevant since the hippocampal formation is also believed to play a central role in navigation. Theories on spatial navigation usually build on different mechanisms for localization (registering the current location), mapping (defining the topological or metric relation between different locations) and planning (designing a path between the current and a goal location). Path integration is thought to be a fundamental part of any biologically realistic spatial navigation model, see, e.g., McNaughton, Battaglia, Jensen, Moser, and Moser (2006). Regarding the hippocampus, place cells are traditionally thought to realize path integration, but as Kubie and Fenton (2009) claim, head direction cells alone may be able to solve this task. In correspondence to these assumed functions, our proposal about estimating the sensory signals as ARMA processes with independent driving sources resulted in a model which is able to (1) extract the hidden driving sources (identified as events), (2) represent separately the hidden states and (3) learn a transformed form of the hidden dynamics that allows for prediction of relevant changes (events) in the environment even if full access to the observation cannot be granted. Here we argue that the resulting internal model is also a natural placeholder for path integration. In turn, our model can be seen as a link between episodic memory and spatial navigation. In the following let us elaborate on this notion.