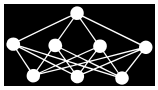

TOWARDS A UNIFIED MODEL OF CORTICAL COMPUTATION II:

FROM CONTROL ARCHITECTURE TO A MODEL OF CONSCIOUSNESS

A. Lőrincz *

Abstract:

The recently introduced Static and Dynamic State (SDS) Feedback control scheme together with its modified form, the Data Compression and Reconstruction (DCR) architecture that performs pseudoinverse computation, suggests a unified model of cortical processing including consciousness. The constraints of the model are outlined here and the features of the cortical architecture that are suggested and sometimes dictated by these constraints are listed. Constraints are imposed on cortical layers, e.g., (1) the model prescribes a connectivity substructure that is shown to fit the main properties of the 'basic neural circuit' of the cerebral cortex (Shepherd and Koch [1], Douglas and Martin [2] In: The synaptic organization of the brain, Oxford University Press, 1990), and (2) the stability requirements of the pseudoinverse method offer an explanation for the columnar organization of the cortex. Constraints are also imposed on the hierarchy of cortical areas, e.g., the proposed control architecture requires computations of the control variables belonging to both the 'desired' and the 'experienced' moves as well as a 'sign-proper' separation of feedback channels that fit known properties of the basal ganglia – thalamocortical loops [3]. An outline is given as to how the DCR scheme can be extended towards a model for consciousness that can deal with the 'homunculus fallacy' by resolving the fallacy and saving the homunculus as an inherited and learnt partially ordered list of preferences.

Key words: *Neural network, cortical computation, control architecture, consciousness*

Received: December 22, 1996
Revised and accepted: January 23, 1997

*Lőrincz
Department of Photophysics of the Institute of Isotopes
Hungarian Academy of Sciences, Budapest, P.O.Box 77, Hungary H-1525
Department of Adaptive Systems
Attila József University, Szeged, Dóm square 9, Hungary H-6720

1. Introduction

Growing interest is still very much in evidence concerning the construction of models on cortical processing. The explosion of knowledge in the neurosciences and the advent of computational methods using parallel architectures made of simple adapting linear and/or nonlinear elements have given rise to the flourishing of new ideas. Of late, seeking of a neuronal substrate for consciousness has also become a primary research issue [4, 5] and this shift was boosted by (1) new results on visual awareness [6]: that normal observers in a forced choice situation can reliably locate targets they seem to be unaware of, and by (2) the discovery concerning perception of bistable images that some of the neurons in the visual processing stream, possibly in the deeper fifth and sixth layers of the cortex (V1/V2, V4 and MT), show correlated firing with the subject's subjective perceived state even if the visual scene does not change [7, 8][1].

The main point of this paper is to show that the control architecture described previously [9, 10, 11, 3] and its modified form that solves overdetermined matrix equations and is equivalent to Wittmeyer's iterative scheme [12] suggested for data compression and data reconstruction may be further simplified giving rise to a layered structure having some resemblance to the basic neural circuit of the neocortex [1, 2]. Taking this observation as our starting point we will analyze whether the computational scheme could be viewed as a general model of the cortical architecture. In that the overall task is tremendous, obviously we cannot deal with the vast literature and the broad range of findings related to the subject. Instead, the paper follows the inherent logic of the architecture and provides a construct that seems rather promising since, without any further assumption, one has (1) a model of the cortical layers, (2) the mathematical need for the cortical columnar structure, (3) a control architecture that can control plants of any order [13] giving rise to architectural constraints that resemble the division of the motor processing areas, and (4) a starting point for the model of consciousness that has no homunculus fallacy: the view that once an internal representation is created it is still meaningless unless someone can read it (see, e.g., [14]), and the better the representation – and we believe that our representation closely mirrors the external world – the more elaborate is the necessary reader. So the infinite series of questions is as follows: Where is that reader? What kind of representation is that reader using? Who is making sense of that representation, and so on ad infinitum.

2. From the control architecture to the layered data compression and data reconstruction scheme

The Static and Dynamic State (SDS) Feedback Control architecture for first order plants [11] can be described as follows (see Fig. 1): We have a feedforward controller, an estimate of the inverse dynamics. We "know" the state of our plant, this is our observed state ($\mathbf{q}$, $\mathbf{q} \in \mathbf{R}^n$); our aim is to modify that state with the speed vector $\dot{\mathbf{q}}_\mathbf{d}$, where subscript d denotes 'desired' and the dot stands for tempo-

[1] Change of the subjectively perceived state without changing the visual scene may occur, for example, with the famous Necker cube or, as in the experiments of Leopold and Logothetis [8], by means of binocular rivalry.

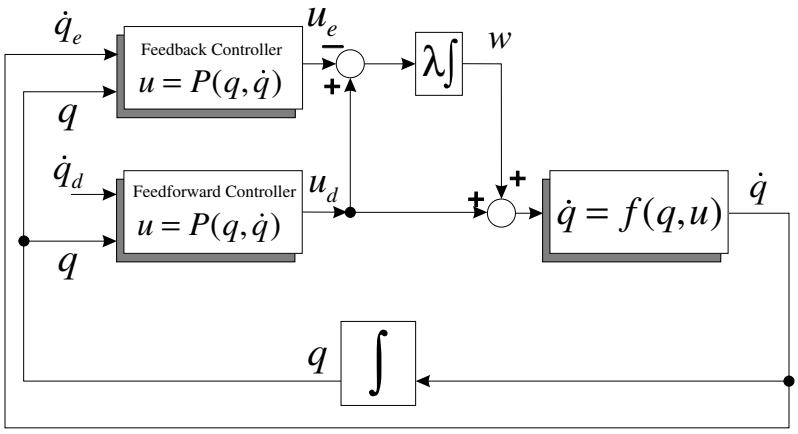

Fig. 1. Architecture of the Static and Dynamic State Feedback Control network. *The same inverse dynamics controller plays two roles: it is used to compute the desired control signal based on the static information about the state and the desired speed of the plant as well as the compensatory signal based on the experienced speed of the plant. These two control signals develop the compensatory signal of the desired control signal.*

ral differentiation. The feedforward controller provides us with the desired control vector $\mathbf{u_d} = \mathbf{p}(\mathbf{q}, \dot{\mathbf{q}}_\mathbf{d})$, where $\mathbf{p}$ denotes the estimate of the inverse dynamics and $\mathbf{u_d} \in \mathbf{R}^N$ with n smaller or equal to N. The control vector $\mathbf{u_d}$ when applied to the plant exhibiting the first order dynamics

$$\dot{\mathbf{q}} = \mathbf{f}(\mathbf{q}, \mathbf{u_d}) \tag{1}$$

should result in the desired speed vector $\dot{\mathbf{q}}_\mathbf{d}$. Our knowledge about state $\mathbf{q}$, or the feedforward controller model of the inverse dynamics itself may be imprecise and we experience a speed vector $\dot{\mathbf{q}}_\mathbf{e}$ (where subscript e denotes 'experienced') and $\dot{\mathbf{q}}_\mathbf{e}$ may differ from $\dot{\mathbf{q}}_\mathbf{d}$. The SDS philosophy is that we should use the very same controller or an identical copy of it to compute a control vector for the observed state with the experienced speed vector and then use this control vector as a means for correction. The SDS scheme thus computes the experienced control vector $\mathbf{u}_\mathrm{e} = \mathbf{p}(\mathbf{q}, \dot{\mathbf{q}}_\mathbf{e})$ and adds the time integrated correction control vector $\mathbf{w}$ to the 'desired' control vector:

$$\begin{aligned} \mathbf{u} &= \mathbf{u_d} + \mathbf{w} \\ \dot{\mathbf{w}} &= \lambda(\mathbf{u_d} - \mathbf{u}_\mathrm{e}) \end{aligned} \tag{2}$$

where $\mathbf{u_d}$ and $\mathbf{u}_\mathrm{e}$ are both computed by the estimate of the inverse dynamics, and λ is the gain factor. It can be shown that the resulting scheme is stable provided that the perturbation is 'sign-proper'. 'Sign-properness' means that for some perturbations a change of sign is needed in the feedback channel and the method requires the recognition of this change and needs a dedicated channel with corrected feedback signs for proper control [9, 11]. A simple example is the case

of mirror writing when without the recognition of the necessary reversal of control the SDS scheme cannot work. It can be shown that the scheme is not restricted to first order plants and, in fact, plants of any order can be controlled with a slightly modified architecture; the only requirement is that the list of variables to be measured should be extended beyond the measurement of the static state variables [13].

Figure 1 depicts the SDS architecture and shows the two controllers (the feedforward controller, 'ff', and the feedback controller, 'fb', being identical copies of each other), the plant and the connection structure of the control architecture, as well as the sensory feedback from the plant to the controller.

The Data Compression and Reconstruction (DCR) network can be developed from the SDS scheme. As is described in the accompanying paper [12], on replacing the feedforward controllers of the inverse dynamics with a memory matrix and the plant by the very same memory matrix in transposed position we achieve a data compression scheme identical to Wittmeyer's iterative scheme for solving matrix equations [15]. Figure 2 depicts the working of the algorithm, which has an input vector $\mathbf{x_0}$ ($\mathbf{x_0} \in \mathbf{R}^N$) and a 'direct internal representation' $\mathbf{a_d}$ ($\mathbf{a_d} \in \mathbf{R}^n$) connected by memory matrix $Q = Q(i,j)$ ($Q \in \mathbf{R}^n \times \mathbf{R}^N$) formed by n memory vectors $\mathbf{q}_i, i = 1, \ldots, n$, or neural units of dimension N. The memory vector (or memory trace) $\mathbf{q}_i$ represents the feedforward connection structure of neuron i that connects the inputs to neuron i. (The respective $\mathbf{q}_i$ vectors are not shown in the figure.) The low dimensional internal representation is then used to reconstruct the input. The reconstructed input is again inputted to an identical copy of the same system to measure the 'goodness' of the internal representation. The output of this identical copy is the 'experienced' internal representations ($\mathbf{a_e}$). The difference between the 'direct internal representation' and the experienced internal representation is used to correct the internal representation in the same way as is done in the SDS model:

$$\begin{aligned} \mathbf{a_d} &= Q^T \mathbf{x_0} \\ \mathbf{y} &= Q\mathbf{a} \\ \mathbf{a_e} &= Q^T \mathbf{y} \\ \mathbf{a} &= \mathbf{a_d} + \mathbf{w} \\ \dot{\mathbf{w}} &= \lambda(\mathbf{a_d} - \mathbf{a_e}) \end{aligned} \tag{3}$$

where $\mathbf{y}$ is the noise filtered reconstructed input, and $\mathbf{a}$ is the relaxing corrected internal representation featuring memory based noise filtering. Noting that

$$\mathbf{a_d} - \mathbf{a_e} = Q^T(\mathbf{x_0} - \mathbf{y}) \tag{4}$$

one can modify Eqs. 3 as follows:

$$\begin{aligned} \mathbf{a} &= Q^T \mathbf{x} + \mathbf{w} \\ \dot{\mathbf{w}} &= Q^T \dot{\mathbf{v}} \\ \dot{\mathbf{v}} &= \lambda(\mathbf{x_0} - \mathbf{y}) \\ \mathbf{y} &= Q\mathbf{a} \end{aligned} \tag{5}$$

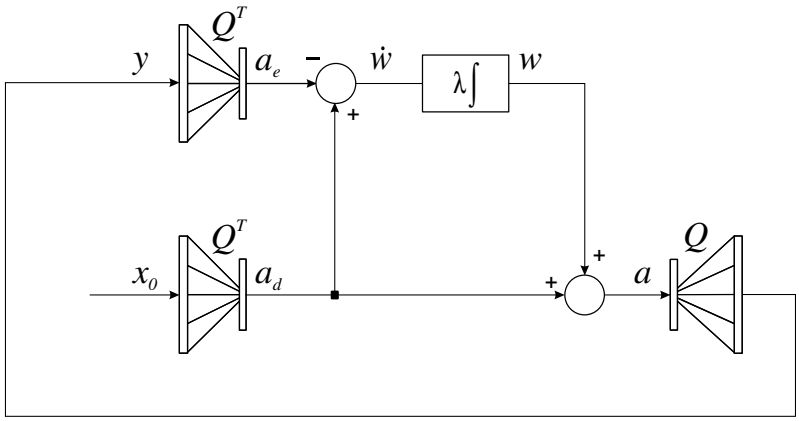

Fig. 2. Architecture of the Dynamic State Feedback data compression and data reconstruction network. *The network is made up of three replicas of the original memory matrix Q: the first replica receives the input vector* $\mathbf{x_0}$ *and gives rise to the internal vector* $\mathbf{a_d}$*; the second utilizes the same matrix in transposed position and computes the reconstructed input* $\mathbf{y}$*; the third deals with the reconstructed image and computes the experienced internal representation* $\mathbf{a_e}$*. The internal representation undergoes (1) differencing between* $\mathbf{a_d}$ *and* $\mathbf{a_e}$*, (2) integration, (3) amplification by* λ*, and (4) addition to form the corrected internal representation* $\mathbf{a}$*.*

These equations lead to a different architecture where, instead of three identical copies of the same memory matrix being used to compute the internal representation, the computation is accomplished in a layered structure (Fig. 3). This architecture is somewhat similar to the routing network of Olshausen, Van Essen and Anderson [16] as the output interacts with the inputs and modifies the effect of the feedforward connections of the neurons. The philosophies and the underlying dynamics of the two networks are, however, different. The routing network dynamically reshapes the feedforward connections in order to match a given memory trace and acts as an invariant representation. The DCR network does not aim to match a memory trace but rather behaves as a noise filter based on its memory content and reconstructs the input using that filtering just as do networks that utilize principal component analysis (PCA).

For the sake of later considerations it is worth noting that supervisory information for input reconstruction can be entered at the output stage of the network (see Fig. 3) via enforcing (biasing) output nodes to take (to approximate) predetermined output values.

In the following the connections and similarities between the DCR architectures and the cortical neuronal architecture will be listed. The similarities are rather broad and may serve as starting points for further research to determine the place of the DCR scheme in cortical modelling.

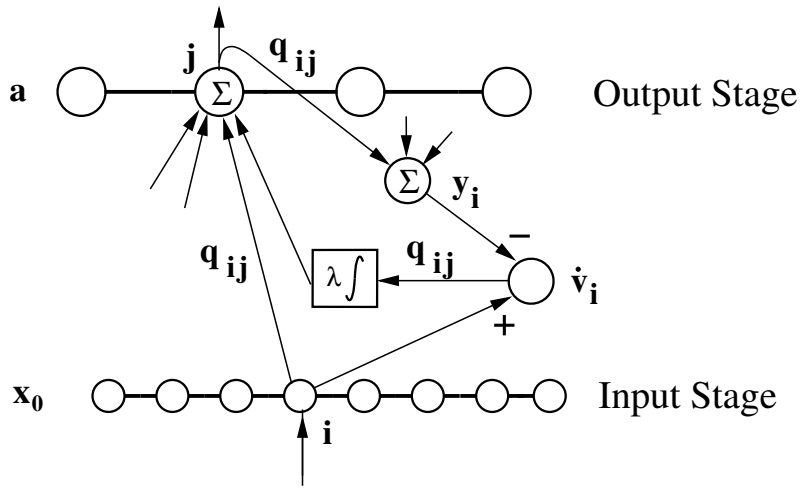

Fig. 3. Layered architecture of the Dynamic State Feedback data compression and data reconstruction scheme. *The input gives rise to two parallel excitations: (1) direct excitation of output (***a***) the connection matrix ($Q(i,j) = q_{ij}$) and (2) excitation of feedback arm (***v***). The output provides inhibitory input to the same feedback channel. The inhibitory input is weighted by the same connection matrix. The feedback channel is integrated by time and serves as the correcting term of the input.*

3. Consequences for cortical intraareal processing

The network sketched in Fig. 3 may be part of information processing in the neocortex. Such processing is rather complex and simplified schemes have been published [1, 2]. The 'basic circuit' of Shepherd and Koch ([1] p. 22.) does not exclude the DCR suggestion: In the 'basic circuit' afferent inputs excite the pyramidal output neurons in two ways; directly and indirectly. In the former case the afferents of the cortical layer excite pyramidal cells directly, whereas in the latter case the afferents first excite the spiny stellate cells of lamina IV and then the spiny stellates excite the pyramidal cells. Also, the output of the pyramidal neurons gives rise to intrinsic recurrent axon collaterals that excite smooth stellates of inhibitory nature, in turn giving rise to negative feedback to the same pyramidal neuron as well as to neighbouring pyramidal neurons. As a first approximation we accept the widely held view that neurons are 'integrate and fire' devices and that merely the frequency of neural spiking represents the main body of the signal transmitted by the neuron. Within this approximation the indirect excitation and the indirect inhibition of pyramidal neurons may be considered as excitation and inhibition *with* (leaky) integration when compared with the direct excitation of the same pyramidal neurons that has no integration in it. Now, the components of the network of Fig. 3 can be identified with the 'basic network' by noting that differencing and integration are interchanged in their computational order in the two architectures.

The consequences for the DCR model are rather strict. The first consequence

of the DCR scheme is the requirement that both the output and the reconstructed input should utilize the same matrix. In other words if input component A excites output neuron B via connection C both in a direct and an indirect fashion, and if connection C has connection strength D then the output of neuron B should give rise to an inhibitory action to the input side of neuron B and the connection strength of this inhibitory component should be proportional to D. Also, the inhibitory action of neuron B should be fed back to the input side of neuron B for *each* input separately. These features may be considered as the predictions of the model and may be given a yes or no answer by single cell level recording experiments.

The second consequence of the DCR scheme originates from the pseudoinverse procedure itself. The pseudoinverse procedure becomes somewhat unstable if very similar memory traces are allowed to develop. In high dimensional input spaces it is hard to develop quasi-orthogonal memories with large networks featuring Hebbian tunability. It seems reasonable to restrict the high dimensional input spaces to smaller dimensional ones and to use a hierarchy of DCR layers together with a stepwise procedure for the forming of the memory traces. This property may explain the need for developing the topographical structure of minicolumns containing approximately 100 neurons [17].

The third consequence is that tuning of the network can give rise to memory matrix Q being a projection matrix by itself. An example is if the tuning is led by input statistics, in which case a projection matrix may be achieved. As soon as the memory matrix approximates a projection matrix reasonably well the DCR scheme becomes a feedforward architecture since the feedback channel will not give rise to any correction. This feature is in line with the observation of Oram and Perrett [18] that in spite of the extensive feedback connections within and between layers the speed of visual processing suggests feedforward computation at least up to the inferotemporal cortex. According to the DCR scheme, departures from feedforward computations may be found for novel visual inputs.

The fourth consequence is the following: if differencing is not perfect, then the result of the computation is a rescaled reconstructed input. In other words if $\mathbf{x_0}-\mathbf{y}$ is replaced by $\mathbf{x_0}-\kappa\mathbf{y}$ then the output will be larger (smaller) if κ is smaller (larger) than 1. This normalization may be speculated as serving dynamic adaptation to signal levels that retain the dynamic range for signal transfer while allowing responses in a broad range of input intensities. In fact, taking into account the suggestion of Heeger, Simoncelli and Movshon [19] that motion detection undergoes normalization at both the V1 and MT levels, the normalizing option κ may be a general ingredient and may be directly formed by other parts of the neuronal circuit.

4. Consequences for cortical interareal processing

Most of the consequences listed below concern the motor areas. The SDS scheme has been suggested as a suitable candidate for constructing a model of the high order motor functions of the basal ganglia – thalamocortical loops [3]. The SDS model of motor control predicts the existence of cortical neurons that compute the control actions of the 'desired motion' and others that compute the control actions for the 'experienced motion'. The first type of neurons can be interpreted as neurons that show preparatory activity before motion starts. The second type

of neurons show no preparatory activity and start to fire at the onset of motion. Neurons featuring these properties have been found in these loops [20, 21].

Another feature of the SDS model is the differencing accomplished between 'desired' and 'experienced' channels. These two channels may be identified with the direct and indirect pathways of the basal ganglia [3] since there appears to be a functional consistency between the various projections to the external and internal segments of the globus pallidus [22]. Cortical afferents arising from the motor areas target the medium spiny neurons of the putamen. Activation of striatal medium spiny neurons associated with the different arms of the indirect pathway will tend to increase the output of the basal ganglia. In contrast, activation of medium spiny neurons associated with the direct pathway tend to decrease the output of the basal ganglia. The overall result is that cortically initiated activation of the direct (indirect) pathway will tend to enhance (suppress) reentrant thalamocortical excitation by decreased (increased) inhibitory outflow from basal ganglia to the thalamus [22] thereby allowing one to identify the direct (indirect) pathway with the 'desired' ('experienced') channels of the SDS scheme.

The finding that the basal ganglia – thalamocortical loops are organized in distinct parallel pathways [23, 24, 25, 20, 21, 26, 27, 28, 22] is then suggested as the division of the 'task space' into subsets with feedback organized into sign-proper channels in accord with the SDS scheme. A detailed description of similarities between the basal ganglia – thalamocortical loops and the SDS scheme can be found elsewhere [3].

5. Extending the DCR scheme to the temporal domain

The DCR network provides a framework for the reconstruction of sensory inputs. The sensory inputs, however, may change rather quickly and the question arises how to enter the time domain and create a new (complementary) set of 'sensors' that could be used to reconstruct inputs with time dependency, or in other words, to reconstruct the input as well as the phenomenon that the input is subjected to. The issue is somewhat similar to one of the problems in computer vision: one has a set of snapshots of the visual image and would like to know the subsequent images. The solution is in the development of the flow field belonging to the actual motion(s). The flow field is a vector field and provides information about the change of the input in terms of the direction of motion at each point and at every instant. The flow field allows one to forecast the visual scene in the immediate future. The problem is general and one can ask how to go beyond this problem of image processing to develop the 'generalized flow field' belonging to any given sensory layer.

Efforts to develop ANN models of the generalized flow field have been made in the literature. One means of describing the generalized flow field is to consider it as a vector field given in terms of the flow-field sensors, a set of secondary sensors that are inputted by the original sensors of a given layer. These secondary sensors can be set up from directed connections bridging the original sensors. The directed connections measure the component of the flow along the connection and can play a predictive role since boosting the flow will advance the sensory input(s) by time (see e.g., [29]). Thus the flow field extends the original sensory system

and can be used to forecast the sensory input via predictive dynamic remapping: The flow field can advance the input and this advancement can be fitted to the delays of the processing. If carefully adjusted, this predictive dynamic remapping promotes the input reconstruction in spite of the various delays in the processing. The connection structure that describes the flow field and thus the structure of temporal development will also be called temporal association.

Since temporal associations are the means of prediction, this point of the model becomes similar to the predictive top-down Kalman filtering structure of Rao and Ballard [30]. Beyond the dissimilarities of the respective architectures the main difference between the two models is that our starting point is a general control architecture that can be modified to a compression–reconstruction model and then to a layered structure that resembles the 'basic circuit' of cortical processing whereas the predictive top-down Kalman filtering model is a computational construct that deals with the problem of sensory processing and promotes memory based noise filtering (just like the DCR scheme) but without input reconstruction (unlike the DCR scheme). At the same time the top-down predictive Kalman filtering may be the solution to form generalized flow fields (i.e., temporal associations) and thus may be used to fill the missing point of the DCR architecture.

Another way of establishing temporal associations is as follows. (1) Self-organizing means are capable of developing an 'approximate geometry representation' [31, 32]. (2) The secondary sensors can be built onto the geometry representing directed connections that bridge the local approximator units [33, 10, 11]. (3) The secondary sensors related to the neighbour connections are appropriate devices for measuring the projection of the local flow vector onto the unit vector directed along the connection [34]. It remains an open question how to set up an appropriate local measuring circuit to each connection but it has been shown that the necessary information can be learnt, e.g., by delayed processing [35, 36], thus the neighbour relation representing connection structure can form temporal associations. It may be worth noting that the DCR scheme is a suitable substrate for delayed associations owing to the delayed reconstruction of the inputs. It is possible though, that solutions relying on the problem of reconstruction offer other alternatives.

Possibly the first attempt towards unifying the two routes, i.e., using reduced dimensional representation followed by temporal association to form one layer of a hierarchical information processing system, was made by Neven and Aertsen [37] and was used to model the formation of receptive fields in the early processing areas. The suggested hierarchical processing scheme has lately been given support at a higher cortical level, i.e., at the level of perirhinal/entorhinal – hippocampal formation [38].

6. Modeling consciousness with DCR, temporal association, and binding

It is fairly straightforward to construct a model of conscious awareness based on the architecture of Fig. 3: (1) The homunculus, that is the system making use of the internal representation (not modelled by the DCR scheme), is the functional system of inherited/learnt and partially ordered list of preferences. (2) The input is processed via a bottom-up network and the connected top-down network. (3) The

outputs of that network are the inputs of the homunculus. (4) The homunculus plays a supervisory role on the outputs of the bottom-up network by biasing or determining some or all of the components of the output and thus the homunculus influences the dynamic *input* reconstruction process. Without the homunculus the network is a simple input reconstruction (or noise filtering) machine, while the supervisory role of the homunculus modifies the reconstruction procedure and one may say that the reconstruction is subject to a 'frame system' [39, 40], the preference system of the homunculus.

In terms of the homunculus fallacy: the homunculus is the functional system of inherited and learnt preferences. Preferences relate to tasks and these tasks can be ordered according to the preferences. Thus the homunculus includes a set of possible and/or actual tasks. Each task (the description of the task, the order of the control variables that can/may accomplish the task) corresponds to some internal neural sets that can be activated by a dedicated sensor or sensor sets. The preference list is a partially ordered list with importance measure: the higher the importance of an active task the earlier that task will be accomplished or will be given a try to accomplish it. At each instant the homunculus is engaged in accomplishing one of its own tasks selected according to its own preferences. Task accomplishment is an inherited or learnt action or action series that deactivates or tries to deactivate the dedicated sensor(s) of the actual task selected by the preference list. A scheme of this sort equipped with learning capabilities can be formulated in terms of reinforcement learning by giving an event based description of the problem [41, 42, 43]. It has been shown that such event based formulation allows generalization, i.e., higher order concept formation [44].

Since some of the activities of the homunculus correspond to the outputs of the sensory (DCR) processing and these DCR outputs are supervised by the homunculus in part (the homunculus can influence these outputs and the outputs are the common result of the homunculus's activity and the DCR processing) or in full (the result of the DCR processing and thus the input reconstruction too are determined solely by the activity of the homunculus) one may say that in the DCR scheme input reconstruction is a sensory input initiated homunculus mediated and supervised process. The key point of this architecture is that the homunculus contains a preference list but this list is not the homunculus' internal representation of the world. The homunculus' representation of the world is outside the homunculus and is nothing but the memory filtered, homunculus mediated, and DCR reconstructed sensory input. This proposition thus shortcuts the homunculus fallacy. Also, the proposition forms a model of conscious awareness: the homunculus mediated and reconstructed sensory input, this being the dynamic frame system of the homunculus, i.e., the homunculus' actual representation of the world, forms the actual conscious experience.

It was Mangan who first suggested that there are two networks in the brain, a bottom-up network equipped with attractor dynamics, and a second network that generates a goodness-of-fit metric about the working of the basic network [45]. This view is somewhat modified by the DCR scheme in that the DCR scheme assumes three networks: (1) the bottom-up basic network, (2) the network within the homunculus (e.g., the preference list and task representation), and (3) the homunculus modulated top-down network that together with the bottom-up network

creates a goodness-of-fit metric in terms of the goodness of the reconstructed input itself.

The DCR scheme can be read as bottom-up processing that coexists with memory based top-down expectations having considerable resemblance to Grossberg's Adaptive Resonance Theory (ART) (see e.g., [46]). The conceptual difference between ART and DCR is that the latter works on the reconstruction problem and thus the reconstructed input can be considered as the homunculus' representation of the sensory inputs.

This simple model of conscious awareness, however, cannot stand as it is since the input is changing and due to delays in the processing the reconstruction is just not possible. If we want to keep this model of conscious awareness then we should deal with the problems imposed by the delays and face the input reconstruction problem in the presence of delays.

Electrophysiological measurements have revealed the correlated firing of neurons at different levels of cortical processing [47, 48, 49, 50]. This correlated activity is called binding [51] and has been suggested as playing an important role in cortical processing, namely, it is thought that conscious experience manifests itself in correlated firing [51, 52, 53]. We shall show below that the same thought emerges in a natural fashion within the DCR scheme.

Our starting point is that the cortical architecture is a DCR hierarchy constructed from DCR layers. The DCR layers may form DCR subhierarchies and these subhierarchies may overlap. Different processing channels, e.g., information coming from different modalities or through different information processing channels using the same sensory system, give rise to different delays within the different DCR layers placed at different points of the DCR hierarchy. It is assumed here that the correcting generalized flow fields will compensate these differences and lock the firing by means of the sensory information and via the constraint of reconstruction at each sensory layer where now every input layer of a DCR layer may play the role of a sensory layer subject to temporal reconstruction constraints. Each DCR layer (or a subhierarchy of DCR layers as well as the full DCR hierarchy) can be the subject of reconstruction if equipped with the appropriate reconstruction tool, the generalized flow field sensory system belonging to that substructure. In other words, the generalized flow field is a suitable general tool to achieve the simultaneity of information content at certain levels of the hierarchical processing in spite of the delays in the processing itself. The setting of the correcting flow field may be considered as a binding procedure that connects different sensory information sources to each other and results in the 'global image' [51, 52, 53]. Global image formation is a natural consequence of the input reconstruction task in the DCR scheme. In other words, the DCR scheme when extended by an appropriate model for temporal association can be thought of as a continuous global image formation (or binding) machine.

Binding thus provides us with a continuous stream of global images and the continuously bound images form a 'global movie': this global movie can serve further temporal associations, now temporal associations can be developed between parts of these global movies. The property that separates these 'global cuts' can arise in connection with changing task reorderings. In other words, a global cut, that we will call a phenomenon can be labeled by the expected associated task order

– that we call expected behaviour. The expected task order is determined by the homunculus, and temporal associations can be made between differing task orders that follow each other. A task order becomes obsolete and a new task order is set when the first task of the expected series has been accomplished or when a new task suddenly emerges, etc. In this sense temporal associations beyond global image formations involve the homunculus. The experienced and learnt time ordering of task accomplishment can serve reasoning at a later stage: Temporal association now says that an actual global cut, or phenomenon X will probably (typically) be followed by future phenomenon Y and/or future phenomenon Z, etc., where the phenomena are labeled according to the task structure of the homunculus. These phenomena, if remembered then form the (behaviourally important and memorized) event system of the homunculus, the basis of a reasoning scheme that may be used to reorganize or upgrade the inherited/learnt preference list [41, 42, 44]. It has been emphasized that "Conscious thought is comparatively slow, serial, and abstract; it deals with only a few objects at a time; its contents are readily translated into a communicable form (i.e., language); and its storage and processing limits can be overcome by the use of external objects such as books, calculators, maps, and word-processing programs" [5]. According to the arguments given above the DCR scheme extended with temporal association is a promising candidate for constructing a model of 'conscious thought' via the interaction between learnt preferences and conscious experience, i.e., via input reconstruction. The temporal association that serves as a learning tool to modify the existing preference list can be considered as a relational architecture such as, for example, the Dynamic Concept Model [42, 44]. The importance of constructing relational architectures when modelling consciousness has been stressed by Taylor [54, 55].

Let us step back and consider the problem of global image formation. Let us assume that temporal association is accomplished somehow. We have taken the view that the temporally correct reconstructed inputs of the different levels of the processing hierarchy that fulfil the top-down constraints imposed by the homunculus form our actual conscious experience: It is the generalized flow field that can be adjusted in a dynamical fashion in order to establish a coherently firing subset of neurons at all the different stages of the computational hierarchy where coherence means that top-down and bottom-up processing and the associated generalized flow-field compensate the computational delays of the system for some subset of the sensory input. The dynamic adjustment of the flow fields then allows 'backward referral in time', a unique property found by Libet et al. [56, 57] when studying event readiness potentials of the motor areas and their conscious correlates.

7. Conclusions

In the present work a model for cortical computations has been formulated. The model utilizes a dynamic state feedback structure that can be used both for control [9, 10, 11] and for data compression and data reconstruction [12]. The controlling Static and Dynamic State (SDS) scheme predicts a control architecture that closely resembles the basal ganglia – thalamocortical loops. Also, the modified SDS architecture, the Data Compression and Reconstruction (DCR) scheme can be reduced to a layered structure that resembles part of the 'basic network' [1, 2] of cortical

layers.

It has been shown that the DCR scheme is a suitable substrate for temporal associations because of the delayed reconstruction of the inputs. Further, a possible method for constructing temporal associations, i.e., generalized flow fields has been described. The DCR scheme and the temporal association together can give rise to global image formation, phenomenon reconstruction, and prediction. The DCR model of conscious awareness claims that the reconstructed inputs of all DCR layers with the reconstruction being mediated by the partially ordered list of inherited and learnt preferences forms the actual conscious experience.

The scheme shortcuts the so called homunculus fallacy by saying that the homunculus, i.e. the system that makes use of the internal representation, is the functional system of inherited and learnt preferences and this homunculus influences the DCR reconstruction procedure by influencing the activities of the internal representation. The homunculus' representation of the world is the reconstructed input that has been mediated by the homunculus and is outside of the homunculus. The DCR model of consciousness can be viewed as the dynamic frame system of the homunculus.

The reconstruction is constrained by the sensory inputs, since the delays of the processing should be compensated. Generalized flow fields have been suggested as the flexible means of delay compensation. Delay compensation then allows locking of neuronal spiking and it is argued that the homunculus mediated reconstructed input and the rest of the input that is not reconstructed by the supervisory actions of the homunculus should coexist and thus we identified the neuronal subsets exhibiting locked firing - that are locked by parts of the neuronal correlates of the functional system of the homunculus - with the actual conscious experience. It should be mentioned though, that locked firing may be exhibited by subsets of neurons of the DCR layers and thus the simple fact that some of the neurons exhibit locked firing does not necessarily mean that these neurons are part of the conscious experience. The locked subset of spiking neurons forms the global representation ('global image') of the homunculus of the world.

One can construct a DCR hierarchy of DCR layers and to create a 'global movie' by means of the generalized flow fields subserving the input reconstructing binding procedure. The global movie can be broken down into 'global cuts' by means of labeling with the actual task ordering (behaviour). The task order labeled disjunct global cuts can be used to construct a relational architecture that subserves reasoning, reevaluates preferences and can be considered as a model of conscious thought. Since sensory input is supervised by a top-down architecture the conscious thought can create imageries.

Of particular interest from the point of view of self-supervised formation of temporal associations is the hippocampal formation as (1) it is known to be crucial in forming conscious (declarative) memories (see e.g. [38] and references therein), (2) it includes the subiculum and thus it is strongly influenced by the limbic system, (3) the hippocampal formation is the top area of sensory processing where the problem of temporal association and binding should be emphasized, and (4) it is thought that the medial paralimbic system (that includes the supplementary motor area as well as the anterior cingulate cortex and develops the elaborated basal ganglia – thalamocortical loops) originates from the hippocampal cortex [58, 59].

Points (3) and (4) above indicate that the hippocampal formation may be an intermediate anatomical structure between the two related computational architectures modelled by the SDS scheme (the hierarchy of higher order motor function) and the DCR scheme (the layered structure of cortical processing). An intriguing speculation that arises is that the eventual clue of temporal association may be hidden in the hippocampal formation and may assume some combined form of the SDS and DCR processing schemes.

8. Acknowledgments

I am grateful to Mr. Péter Aszalos and Dr. Gyula Kovács for enlightening discussions. This work was partially supported by OTKA grants T14566 and T17100 and by grant JF No. 519 of the US-Hungarian Joint Fund.

References

[1] G. M. Shepherd and Christof Koch. Introduction to synaptic circuits. In *The synaptic organization of the brain*, pages 3–31. Oxford University Press, New York, 1990.

[2] R. J. Douglas and K.A.C. Martin. Neocortex. In *The synaptic organization of the brain*, pages 389–438. Oxford University Press, New York, 1990.

[3] A. Lőrincz. Neurocontrol III: Temporal differencing models of the basal ganglia – thalamocortical loops. *Neural Network World*, 1997. In press.

[4] F. Crick. *The astonishing hypothesis.* Charles Scribner's Sons; New York, 1994.

[5] I.B. Farber and P.S. Churchland. Consciousness and the neurosciences: Philosophical and theoretical issues. In *The cognitive neurosciences*, pages 1295–1306. Bradford Books / MIT Press, Cambridge, MA, 1995.

[6] F.C. Kolb and J Braun. Blindsight in normal observers. *Nature*, 377:336–339, 1995.

[7] N. Logothetis and J. Schall. Neuronal correlates of subjective visual perception. *Science*, 245:761–763, 1989.

[8] D. Leopold and N. Logothetis. Activity-changes in early visual cortex reflect monkey's percept during binocular rivalry. *Nature*, 379:549–553, 1996.

[9] Cs. Szepesvári and A. Lőrincz. Inverse dynamics controllers for robust control: Consequences for neurocontrollers. In *Proceedings of International Conference on Artificial Neural Networks*, pages 697–702, Bochum, Germany, 1996. Springer-Verlag, Berlin.

[10] Cs. Szepesvári and A. Lőrincz. Neurocontrol I: Self-organizing speed–field tracking. *Neural Network World*, 6:875–896, 1996.

[11] Cs. Szepesvári and A. Lőrincz. Neurocontrol II: High precision control achieved using approximate inverse dynamics models. *Neural Network World*, 6:897–920, 1996.

[12] T. Fomin, J. Körmendy-Rácz, and A. Lőrincz. Towards a unified model of cortical computation I. Data compression and data reconstruction using dynamic state feedback. *Neural Network World*, 1997. In press.

[13] Cs. Szepesvári and A. Lőrincz. Static and dynamic state feedback control of higher order plants. 1997. Submitted.

[14] J.R. Searle. *The rediscovery of the mind.* Bradford Books / MIT Press, Cambridge, MA, 1992.

[15] H. Wittmeyer. Ueber die Loesung von linearen Gleichungssystemen durch Iteration. *Z. Angew. Mat. Mech.*, 16:301–310, 1936.

[16] B.A. Olshausen, D.C. Van Essen, and C.H. Anderson. A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. *Journal of Neuroscience*, 13:4700–4719, 1993.

[17] A. Peters and E. Yilmaz. Neuronal organization in area 17 of cat visual cortex. *Cerebral Cortex*, 3:49–68, 1993.

[18] M.W. Oram and D.I. Perrett. Modelling visual recognition from neurobiological constraints. *Neural Networks*, 7:945–972, 1994.

[19] D. J. Heeger, E. P. Simoncelli, and J. A. Movshon. Computational models of cortical visual processing. *Proc. Natl. Acad. Sci., USA*, 93:623–627, 1996.

[20] G. E. Alexander and M. D. Crutcher. Preparation for movement: Neural representations of intended direction in three motor areas of the monkey. *Journal of Neurophysiology*, 64:133–150, 1990.

[21] M. D. Crutcher and G. E. Alexander. Movement-related neuronal activity selectively coding either direction or muscle pattern in three motor areas of the monkey. *Journal of Neurophysiology*, 64:151–163, 1990.

[22] Garret E. Alexander. Basal ganglia. In *The Handbook of Neural Theory and Neural Networks*, pages 139–144. Bradford Books / MIT Press, Cambridge, MA, 1995.

[23] G. Percheron, G. Yelnik, and C. Francois. In *The Basal Ganglia*, pages 87–105. Plenum Press, New York, 1984.

[24] G. Percheron, G. Yelnik, and C. Francois. A Golgi analysis of the primate globus pallidus: III. Spatial organization of the striato-pallidal complex. *Journal of Neural Computation*, 227:214–227, 1984.

[25] G. E. Alexander, M. D. Crutcher, and M. R. DeLong. Basal ganglia–thalamocortical circuits: Parallel substrates for motor, oculomotor, "prefrontal" and "limbic" functions. *Progress in Brain Research*, 85:119–146, 1990.

[26] G. E. Alexander and M. D. Crutcher. Neural representations of the target (goal) of visually guided arm movements in three motor areas of the monkey. *Journal of Neurophysiology*, 64:164–178, 1990.

[27] G. Percheron and M. Filion. Parallel processing in the basal ganglia: Up to a point. *Trends in Neuroscience*, 14:55–56, 1991.

[28] J. E. Hoover and P. L. Strick. Multiple output channels in the basal ganglia. *Science*, 259:819–821, 1993.

[29] J. Droulez and A. Berthoz. A neural network model of sensoritopic maps with predictive short-term memory properties. *Proc. Natl. Acad. Sci., USA*, 88:9653–9657, 1991.

[30] R. P. N. Rao and D. H. Ballard. Kalman filter model of the visual cortex. *Neural Computation*, 1997. In press.

[31] Cs. Szepesvári, L. Balázs, and A. Lőrincz. Topology learning solved by extended objects: a neural network model. *Neural Computation*, 6:439–456, 1994.

[32] Cs. Szepesvári and A. Lőrincz. Approximate geometry representations and sensory fusion. *Neurocomputing*, 12:267–287, 1996.

[33] T. Fomin, Cs. Szepesvári, and A. Lőrincz. Self-organizing neurocontrol. In *Proceedings of IEEE International Conference on Neural Networks. IEEE World Congress on Computational Intelligence, Orlando, Florida*, pages 2777–2780, Orlando, Florida, 1994. IEEE Publications.

[34] T. Rozgonyi, Zs. Kalmár, and A. Lőrincz. Population coded vector fields and representation of partial differential equations on self-organizing geometry representing artificial neural networks. *International Journal of Neural Systems*, 1997. Submitted.

[35] D. Kleinfield. Sequential state generation by model neural networks. *Proc. Natl. Acad Sci., USA*, 83:9469–9473, 1986.

[36] H. Sompolinsky and I. Kanter. Temporal association in asymmetric neural networks. *Phys. Rev. Lett.*, 57:2861–2864, 1986.

[37] H. Neven and A. Aertsen. Rate coherence and event coherence in the visual cortex: a neuronal model of object recognition. *Biological Cybernetics*, 67:309–322, 1992.

[38] H. Eichenbaum, T. Otto, and N.J. Cohen. Two component functions of the hippocampal memory system. *Behavioral and Brain Sciences*, 17:449–472, 1994.

[39] M. Minsky. A framework for representing knowledge. In *The psychology of computer vision*, pages 211–277. McGraw-Hill, New York, 1975.

[40] K.S. Jones. Frame. In *Catalogue of artificial intelligence techniques*, pages 57–58. Springer-Verlag, Berlin, 1990.

[41] Cs. Szepesvári and A. Lőrincz. Integration of artificial neural networks and dynamic concepts to an adaptive and self-organizing agent. In *Proceedings of International Conference on Artificial Neural Networks*, pages 331–336, Amsterdam, 1993. Springer-Verlag, London.

[42] Cs. Szepesvári and A. Lőrincz. Behavior of adaptive self-organizing autonomous agent working with cues and competing concepts. *Adaptive Behavior*, 2:131–160, 1994.

[43] Cs. Szepesvári. Dynamic Concept Model learns optimal policies. In *Proceedings of IEEE International Conference on Neural Networks. IEEE World Congress on Computational Intelligence*, pages 1738–1742, Orlando, Florida, 1994. IEEE Publications.

[44] Zs. Kalmár, Cs. Szepesvári, and A. Lőrincz. Generalized Dynamic Concept Model as a route to construct adaptive autonomous agents. *Neural Network World*, 5:353–360, 1995.

[45] B. Mangan. Taking phenomenology seriously: The "fringe" and its implications for cognitive research. *Conscious. Cognit.*, 2:89–108, 1993.

[46] G. A. Carpenter and S. Grossberg. A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics and Image Processing*, 37:54–115, 1987.

[47] A.M. Sillito, H.E. Jones, G.L. Gerstein, and D.C. West. Feature-linked synchronization of thalamic relay cell firing induced by feedback from the visual cortex. *Nature*, 369:479–482, 1994.

[48] Z.F. Mainen and T.J. Sejnowski. Reliability of spike timing in neocortical neurons. *Science*, 268:1503–1506, 1995.

[49] A.K. Kreiter and W. Singer. Stimulus-dependent synchronization of neuronal responses in the visual cortex of the awake monkey. *Journal of Neuroscience*, 16:2381–2396, 1996.

[50] M. Abeles. *Corticonics: neural circuits of the cerebral cortex*. Cambridge University Press, Cambridge, UK, 1991.

[51] E. Bienenstock and C. von der Malsburg. Statistical coding and short-term synaptic plasticity: a scheme for knowledge representation in the brain. In *Disordered systems and biological organization*, pages 247–272. Les Houches: Springer-Verlag, Berlin, 1986.

[52] F. Crick and C. Koch. Some reflections on visual awareness. In *Cold Spring Harbor Symp. Quant. Biol.*, volume 55, pages 953–962. 1990.

[53] R. Llinás. Intrinsic electrical properties of mammalian neurons and CNS function. In *Fidia Research Foundation Neuroscience Award Lectures*, volume 4, pages 1–10. Raven Press, New York, 1990.

[54] J.G. Taylor. Can neural networks ever be made to think? *Neural Network World*, 1:4–11, 1991.

[55] J.G. Taylor. Modelling consciousness. *Psyche*, 1996. Submitted.

[56] B. Libet, Jr. E. W. Wright, B. Feinstein, and D.K. Pearl. Subjective referral of the timing for a conscious sensory experience: a functional role for the somatosensory specific projection system in man. *Brain*, 102:191–222, 1979.

[57] B. Libet. Cerebral processes that distinguish conscious experience from unconscious mental functions. In *The principles of design and operation of the brain*, pages 185–202. Springer-Verlag, Berlin, 1990.

[58] F. Sanides. Comparative architectonics of the neocortex of mammals and their evolutionary significance. *Ann. NY Acad. Sci.*, 167:404–423, 1969.

[59] G. Goldberg. Premotor systems, attention to action and behavioural choice. In *Neurobiology of motor programme selection*, pages 225–249. Pergamon Press, Oxford UK, 1992.