Attentional filtering in neocortical areas: A top-down model

András Lőrincz

Eötvös Loránd University, 1117 Budapest, Pázmány Péter sétány 1/D, Hungary

Abstract

Two comparator based rate code models – a reconstruction network model and a control model – are merged. The role of bottom-up filtering is information maximization and noise filtering, whereas top-down control 'paves the way' of context based information prediction that we consider as attentional filtering. Falsifying prediction of the model has gained experimental support recently.

Key words: attention, neocortex, top-down control, reconstruction network

1 Introduction

Here, a comparator model of the neocortex [1] shall be extended by control function in order to model attentional filtering.

2 Model description

The controller. Our control model is formulated in terms of speed-field, that is, state dependent directions pointing towards target positions. A particular speed-field is given, for example, by the difference vectors between the target state and all other states. The control task, called speed-field tracking, is defined as moving according to the speed-field at each state [2]. The dynamic equation of a system is a set of continuous differential equations that determines the change of state per unit time $\dot{\mathbf{x}} \approx \frac{\Delta \mathbf{x}}{\Delta t}$ given the state of the 'plant' $\mathbf{x} \in \mathrm{R}^n$ and the external forces acting upon that plant, including the control action $\mathbf{u} \in \mathrm{R}^p$. Let $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u})$ the dinamics, i.e., what the plant does.

Inverse dynamics works in the opposite way: given the state and the desired change of state, inverse dynamics provides the control vector. Let $\mathbf{v}(\mathbf{x}) \in \mathrm{R}^n$

denote the *desired* speed. Assume that we have an approximate model of the precise inverse dynamics $\mathbf{u}_{ff} : \hat{\mathbf{u}}_{ff}(\mathbf{x}, \mathbf{v}(\mathbf{x})) = \hat{\mathbf{A}}(\mathbf{x})\mathbf{v}(\mathbf{x}) + \hat{\mathbf{b}}(\mathbf{x})$. If $\hat{\mathbf{u}}_{ff}$ influences the plant directly, it is called *feedforward controlling.* For perfect inverse dynamics, the control vector produces the desired speed: $\mathbf{v}(\mathbf{x}) = \mathbf{f}(\mathbf{x}, \mathbf{u}_{ff}(\mathbf{x}, \mathbf{v}(\mathbf{x})))$. If the feedforward control vector is imprecise, then comparison error $\mathbf{e}_c = \mathbf{v}(\mathbf{x}) - \dot{\mathbf{x}}$, the difference between desired and experienced speeds appears. A model of the inverse dynamics can be used to correct this error. The error correcting controller is called *feedback controller* [2] and our model assumes the form $\mathbf{u}_{fb} = \hat{\mathbf{B}}(\mathbf{x})(\mathbf{v}(\mathbf{x}) - \dot{\mathbf{x}})$. The control vector is built from these two terms $\mathbf{u}(t) = \hat{\mathbf{u}}_{ff}(t) + \mathbf{w}(t) = \hat{\mathbf{u}}_{ff}(t) + \int_{-\infty}^{t} \mathbf{u}_{fb}(t')\, dt'$. Our *key assumption* is that the feedforward controller is also a comparator: $\hat{\mathbf{u}}_{ff} = \hat{\mathbf{u}}_{fb} = \hat{\mathbf{A}}(\mathbf{x})(\mathbf{v}(\mathbf{x}) - \dot{\mathbf{x}})$ (Fig. 1). Taken together:

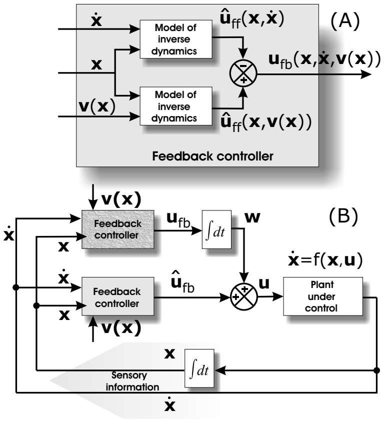

Fig. 1. **Robust controller for speed-field tracking tasks**
A: Model of the inverse dynamics. **B:** SDS feedback controller. (See text.)

$$\mathbf{u} = \hat{\mathbf{u}}_{fb} + \int_{-\infty}^{t} \mathbf{u}_{fb}(t')\, dt' = \hat{\mathbf{A}}(\mathbf{x})(\mathbf{v}(\mathbf{x}) - \dot{\mathbf{x}}) + \int_{-\infty}^{t} \hat{\mathbf{B}}(\mathbf{x})(\mathbf{v}(\mathbf{x}) - \dot{\mathbf{x}}). \qquad (1)$$

Both controllers operate by *comparing* ‘desired’ and ‘experienced’ quantities. This choice makes a useful compromise. It can not be perfect, but explicit modelling of the highly non-linear term $\mathbf{b}(\mathbf{x})$ becomes unnecessary, because it disappears upon differencing. The architecture called the *static and dynamic state* feedback controller, or SDS controller, is globally stable under rather mild conditions [2]: Roughly speaking, transformations are satisfactory if (i) the signs of the components of the control vector and (ii) the domains where these components keep their signs have been determined.

The reconstruction network. The basic reconstruction network (Fig. 2A) has two layers: the reconstruction error layer that computes the difference ($\mathbf{e} \in \mathrm{R}^r$) between input ($\mathbf{x} \in \mathrm{R}^r$) and reconstructed input ($\mathbf{y} \in \mathrm{R}^r$): $\mathbf{e} = \mathbf{x} - \mathbf{y}$ and the hidden layer that holds the hidden representation $\mathbf{h} \in \mathrm{R}^s$

and produces the reconstructed input $\mathbf{y}$ via top-down transformation $\mathbf{Q} \in \mathrm{R}^{r\times s}$. The hidden representation is corrected by the bottom-up transformed form of the reconstruction error $\mathbf{e}$, i.e., by $\mathbf{We}$, where $\mathbf{W} \in \mathrm{R}^{s\times r}$ and is of rank $\min(s,r)$. The process of correction means that the previous value of the hidden representation is to be maintained and the correcting amount needs to be added. In turn, the hidden representation has self-excitatory connections ($\mathbf{M}$), which sustains the activities. For sustained input $\mathbf{x}$ the iteration will stop when $\mathbf{WQh} = \mathbf{Wx}$: The relaxed hidden representation is solely determined by the input and top-down matrix $\mathbf{Q}$. The latter is identified with the long-term memory. BU matrix $\mathbf{W}$ is perfectly tuned if $\mathbf{W} = (\mathbf{Q}^T\mathbf{Q})^{-1}\mathbf{Q}^T$, i.e., if $\mathbf{WQ} = \mathbf{I}$ ($\mathbf{I} \in \mathrm{R}^{s\times s}$). In this case, the network is as fast as a feedforward net.

The network can be extended to support noise filtering by (i) separating the reconstruction error layer and the reconstructed input layer, (ii) adding another extra layer that holds $\mathbf{s} = \mathbf{We}$ and transformation $\mathbf{s} \rightarrow \mathbf{h}$, i.e., $\mathbf{N}$ supports pattern completion [1] and (iii) assuming that BU transformation $\mathbf{W}$ maximizes BU information transfer by minimizing mutual information (MMI) between its components. BU transformed error is passed to the hidden representation layer through transformation matrix $\mathbf{N}$ and corrects hidden representation $\mathbf{h}$ (Fig. 2(B)). MMI plays an important role in noise filtering. There are two different sets of afferents to the MMI layer: one carries the error, whereas the other carries the reconstructed input $\mathbf{y}$ via bottom-up transformation $\mathbf{P}$, followed by a non-linearity that removes noise via thresholding. Thresholding is alike to wavelet denoising, but filters are not necessarily wavelets: They are optimized for the input database experienced by the network. MMI algorithms enable the local estimation and local thresholding of noise components. The method is called sparse code shrinkage (SCS) [3]. Note that SCS concerns the components of the BU transformed reconstructed input: high (low) amplitude components of the BU transformed reconstructed input $\mathbf{Py}$ can (not) open the

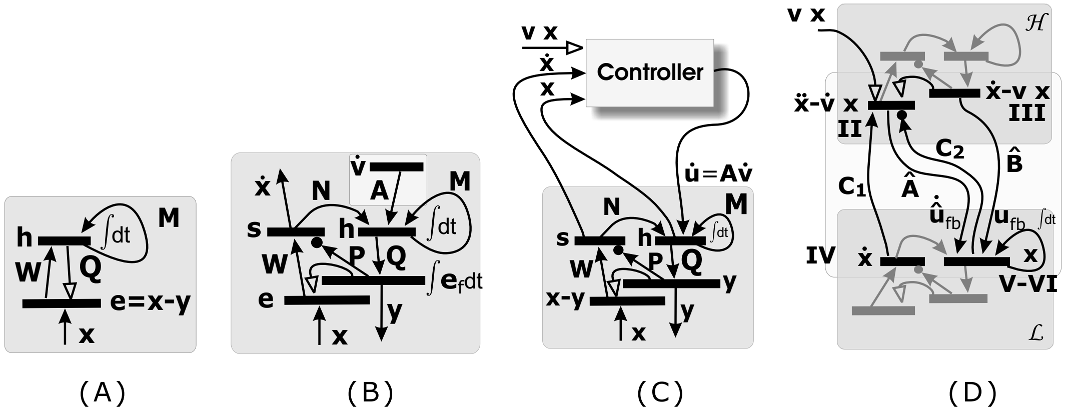

Fig. 2. **Reconstruction networks (A and B) with controllers (C and D)** **A:** Simple reconstruction network (RCN). **B:** RCN with noise filtering. **C:** Controlled RCN. **D:** Controller is an RCN. Roman numbers: mapping to neocortical layers. Arrows with dots: nonlinear modulation. (See text.)

gates of components of the MMI layer and MMI transformed reconstruction error can (not) pass the open (closed) gates to correct the hidden representation. Apart from SCS, the reconstruction network is linear. We shall denote this property by the sign '$\sim$'. '$A \sim B$' means that up to a scaling matrix, quantity A is approximately equal to quantity B. For a *well tuned network* and if matrix $\mathbf{M}$ performs temporal integration, then $\dot{\mathbf{x}} \cong \mathbf{e} \sim \mathbf{s}$ by construction. In a similar vein, hidden representation $\mathbf{h} \sim \mathbf{y}$, apart from noise $\mathbf{y} \sim \mathbf{x}$, and $\mathbf{y} = \int \mathbf{e}_f dt$ where $\mathbf{e}_f$ is the noise filtered version of $\mathbf{e}$.

The joined model. The reconstruction network can be controlled (Fig. 2B,C). Control influences the hidden layer by $\sim \dot{\mathbf{v}}$. Temporal integration occurs at the hidden layer and the effect of the controller is equal to $\mathbf{A}\mathbf{v}$. However, the correct form of matrix $\mathbf{A}$ is not known and the SDS controller is used to achieve approximately perfect control. The controller is made of another reconstruction network $\mathcal{H}$, the 'higher' network, which receives input from the MMI layer of the 'lower', the controlled network $\mathcal{L}$ (Fig. 2D). Control works by subtracting the desired speed from the input of the higher network. The input to network $\mathcal{H}$ is equal to $\dot{\mathbf{x}} - \mathbf{v}(\mathbf{x})$ (via $\mathbf{C}_1$ and from outside, respectively) that can be modulated by $\mathbf{x}$ via $\mathbf{C}_2$. By construction, (i) the input is noise filtered and reconstructed, the reconstructed input is $\sim (\dot{\mathbf{x}} - \mathbf{v}(\mathbf{x}))$ and (ii) apart from the noise content, the error vector approximates the temporal derivative of the reconstructed input. These two differences undergo linear transformations and are added to the hidden representation of network $\mathcal{L}$, where – by construction – they undergo temporal integration. The resulting signals are the control signals needed by the SDS controller provided that control is 'sign proper' [4].

3 Discussion

The perfectly tuned architecture behaves as a bottom-up feedforward network, which is biased by top-down influence. Consider lower reconstruction network $\mathcal{L}$ of Fig. 2(D). The bias will modify the hidden representation of the network $\mathcal{L}$, which may or may not fit the input. If it does not fit, then reconstruction error appears but only a small portion of this error can pass the sparsification process at the MMI layer, because of SCS thresholding. That is, information that matches the *context* of the higher reconstruction network will be able to pass sparsification, whereas other information will be attenuated by the SCS. In turn, top-down influence 'paves the way' of some of the components. This process can be seen as *attentional filtering.*

Now, we shall map the architecture onto the six-layered neocortex. Most prominent neocortical connections are depicted in Fig. 2(D) within the light gray box between networks $\mathcal{H}$ and $\mathcal{L}$. Roman numbers denote neocortical layers. Input arrives at layer IV ($\sim \dot{\mathbf{x}}$). Layer IV neurons send messages to layer

II ($\ddot{\mathbf{x}} - \dot{\mathbf{v}}(\mathbf{x})$) and layer III ($\dot{\mathbf{x}} - \mathbf{v}(\mathbf{x})$). Layer IV neurons send messages also to layer VI $\sim \mathbf{x}$. Superficial neurons provide output down to layer V $\sim \mathbf{x}$ and VI. Layer V provide feedback to layers II and III. The main output to higher cortical layers emerges from layers II and III. The main feedback to lower layers is provided by layer V [5]. The theoretical model and the anatomical structure can be matched by assuming that superficial layers of the lower cortical layer *and* deep layers of the higher cortical layer form *one* functional unit, the reconstruction network [1], representing certain order of the dynamics.

A falsifying prediction of the model concerns the hidden representation layer, which has to maintain its own activities in order to enable additive corrections and temporal integration. Persistent activities in the deep layers but not in the superficial layers of the entorhinal cortex have been found experimentally [6], providing support to our model. Another falsifying prediction of the joined model is that top-down connections between neocortical layers can be interpreted as long-term memories (see also [1]), because these connections are responsible for the relaxed activities of the hidden layers. These connections are generally more numerous than the feedforward connections between the same areas, but the activity flow along these connections is relatively low and suggests a weak functional role [5]. This apparent discrepancy may be resolved by noting that different interpretations may coexist in the brain as it has been made evident in the animal experiments on binocular rivalry [7] and in experiments with several possible visual interpretations [8,9]. If reconstruction concerns a single interpretation, then feedback activity flow should be *small*: There are evidences that activities in V4 (responsible for conscious detection of colors) and V5 (responsible for conscious detection of fast motion) in the monkey are uncorrelated. According to Zeki [10], uncorrelated activities indicate that conscious experiences propagate downwards along parallel channels. Moreover, the conscious binding of the result of the individual conscious experiences seems to be delayed [11]. In turn, it is possible that only one interpretation is communicated downwards at a time.

According to recent measurements, awareness and attention needs to be distinguished [12] and attention increases neuronal activities responsible for the processing of the attended stimuli. Most probably, endogenous attention facilitates the pathways that should be used by the attended stimuli [13]. In our model, facilitation can manifest itself through control action *within* cortical layers. On the other hand, awareness involves recurrent interactions between areas and can be suppressed by backward masking [12]. This recurrent interaction required for awareness is our candidate function of the feedback connections between cortical areas.

Conclusions. A model of neocortical information processing has been presented and mapped onto the neocortex. The model suggests that noise filtering is accomplished by reconstruction networks *between* neocortical layers, whereas top-down control is the task of the cortical layers. The model provides explanations why feedback connections between cortical layers are more numerous than the bottom-up connections between the different areas and why are these connections so quiet. The model allowed us to distinguish between attention and awareness, two delicate and intertwined concepts.

Acknowledgements. This work was supported by OTKA (No. T-32487).

References

[1] A. Lőrincz, B. Szatmáry, G. Szirtes, Mystery of structure and function of sensory processing areas of the neocortex: A resolution, J. Comp. Neurosci. 13 (2002) 187–205.

[2] C. Szepesvári, A. Lőrincz, Applications of Neural Adaptive Control Technology, World Scientific, Singapore, 1997, Ch. Approximate Inverse-Dynamics Based Robust Control Using Static and Dynamic Feedback, pp. 151–179.

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

[4] A. Lőrincz, Controlled hierarchical filtering: Model of neocortical sensory processing, TechRep: http://www.arxiv.org/abs/cs.NE/0308025 (2003).

[5] E. M. Callaway, The mit encyclopedia of cognitive sciences, MIT Press, Cambridge, MA, 2000, Ch. Visual cortex, cell types and connections in, pp. 867–869.

[6] A. V. Egorov, B. N. Hamam, E. Fransén, M. E. Hasselmo, A. A. Alonso, Graded persistent activity in entorhinal cortex neurons, Nature 420 (2002) 173–178.

[7] D. A. Leopold, N. K. Logothetis, Multistable phenomena: Changing views in perception, Trends in Cognitive Sciences 3 (1999) 254–264.

[8] D. A. Leopold, Visual perception: Shaping what we see, Current Biology 13 (2003) R10–R12.

[9] A. Parker, K. Krug, Neuronal mechanisms for the perception of ambiguous stimuli, Current Opinion in Neurobiology 13 (2003) 433–439.

[10] S. Zeki, The disunity of consciousness, Trends in Cognitive Science 7 (2003) 214–218.

[11] A. Bartels, S. Zeki, The temporal limits of binding: Is binding post-conscious?, in: Soc. Neurosci. Abstr., Vol. 11, 2002, p. 260.

[12] V. Lamme, Why visual attention and awareness are different, Trends in Cognitive Sciences 7 (2003) 12–18.

[13] H. Egeth, S. Yantis, Visual attention: Control, representation, and time course, Annual Review of Psychology 48 (1997) 269–297.