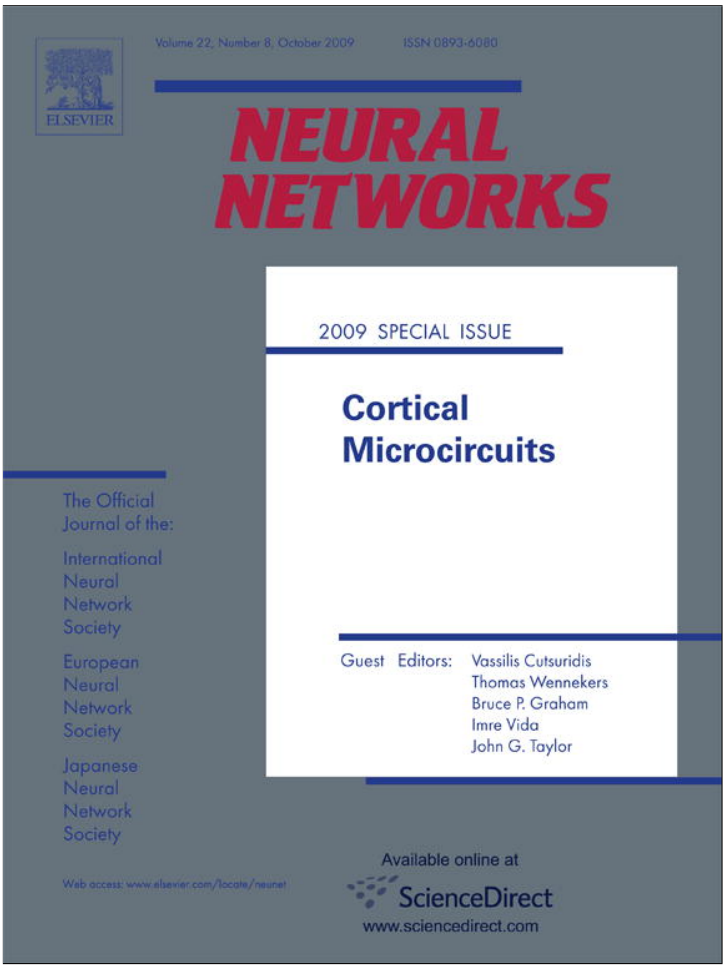
Cortext: A columnar model of bottom-up and top-down processing in the neocortex

Sven Schrader, Marc-Oliver Gewaltig, Ursula Körner, Edgar Körner

2009

Preprint:

This is an accepted article published in Neural Networks. The final authenticated version is available online at: https://doi.org/[DOI not available]



Neural Networks 22 (2009) 1055–1070

Contents lists available at ScienceDirect

Neural Networks

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

2009 Special Issue

Cortext: A columnar model of bottom-up and top-down processing in the neocortex

Sven Schrader *, Marc-Oliver Gewaltig, Ursula Körner, Edgar Körner

Honda Research Institute Europe GmbH, Carl-Legien-Str. 30, D-63073 Offenbach/Main, Germany

ARTICLE INFO

Article history:
Received 18 February 2009
Received in revised form 26 June 2009
Accepted 14 July 2009

Keywords:
Ventral visual pathway
Spike-latency code
Predictive coding
Modulation
Spiking neuron models

ABSTRACT

Experimental data suggests that a first hypothesis about the content of a complex visual scene is available as early as 150 ms after stimulus presentation. Other evidence suggests that recognition in the visual cortex of mammals is a bidirectional, often top-down driven process. Here, we present a spiking neural network model that demonstrates how the cortex can use both strategies: Faced with a new stimulus, the cortex first tries to catch the gist of the scene. The gist is then fed back as *global hypothesis* to influence and redirect further bottom-up processing. We propose that these two modes of processing are carried out in different layers of the cortex. A cortical column may, thus, be primarily defined by the specific connectivity that links neurons in different layers into a functional circuit. Given an input, our model generates an initial hypothesis after only a few milliseconds. The first wave of action potentials traveling up the hierarchy activates representations of features and feature combinations. In most cases, the correct feature representation is activated strongest and precedes all other candidates with millisecond precision. Thus, our model codes the reliability of a response in the relative latency of spikes. In the subsequent refinement stage where high-level activity modulates lower stages, this activation dominance is propagated back, influencing its own afferent activity to establish a unique decision. Thus, top-down influence de-activates representations that have contributed to the initial hypothesis about the current stimulus, comparable to predictive coding. Features that do not match the top-down prediction trigger an error signal that can be the basis for learning new representations.

1. Introduction

Our various brain functions are localized in different regions of the cerebral cortex, the so-called areas. Although cortex areas differ in the details of their cytoarchitecture (Brodmann, 1909), it is generally assumed that they share the same general organization principles.

The cerebral cortex is a thin sheet of tissue with up to six horizontal layers (Peters & Jones, 1984; White, 1989), each with a specific cellular composition and connectivity. Vertically, neurons from all six layers connect to small microcircuits, called columns, which are spread over the cortical sheet (Britten, 1998; Hübener, Shoham, Grinvald, & Bonhoeffer, 1997; Mountcastle, 1997). These compartments are visible in many species, not only anatomically (Lorente de No, 1949; Von Economo & Koskinas, 1925), but also electrophysiologically (Mountcastle, 1957). This has lead researchers to the conclusion that columns are the elementary units of the cortex (Mountcastle, 2003, 1997; Rockland & Ichinohe, 2004).

Columns may differ in their neuron types and numbers, but they all share the same basic connections within and between the cortical layers. Starting from the available anatomical and physiological data about the cortex, we can try to infer which algorithm or function is embodied in the columnar architecture. We can then develop a model that links this architecture to sensory processing.

Clearly, such an endeavor is only possible with additional constraints about the functional and dynamical properties of cortical processing. Additional constraints can be found by looking at how the brain processes sensory stimuli, and we will use vision as our example.

When analyzing a visual scene, our brain faces a combinatorial explosion of possible interpretations. Generally, our sensory environment is complex and ambiguous. Lighting conditions, occlusions, and scene context render a unique interpretation of a given object at least difficult. But even under controlled conditions with well defined contrast and homogeneous lighting conditions, simple visual stimuli lead to a plethora of possible interpretations. These result from the many and often contradicting ways in which low-level features, such as lines and object borders, can be combined to larger features or object parts. And the finer (spatially) the input is analyzed, the more interpretations are possible.

* Corresponding author. Tel.: +49 69 89011 786; fax: +49 69 89011 749.
E-mail address: sven.schrader@honda-ri.de (S. Schrader).

0893-6080/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neunet.2009.07.021

Often, these local ambiguities can only be resolved if knowledge from several distant parts of the scene are taken into account. This suggests that local ambiguities must be resolved by global information which can only be supplied through feedback from higher areas. Such a global hypothesis is also needed to resolve the ambiguities in a realistic scene with occlusions, clutter and inhomogeneous lighting.

Experimental data suggests that a first hypothesis about the content of a complex visual scene is available as early as 150 ms after stimulus presentation (Thorpe, Fize, & Marlot, 1996; Thorpe, 1990). If we consider all involved areas, their response latencies, and also the response properties of cortical neurons, this means that each area cannot contribute more that one spike per neuron. Thus, a single wave of spikes, traveling from the retina through the ventral visual pathway to the highest associative areas must suffice to generate this coarse initial hypothesis about the stimulus.

Based on this reasoning, Delorme and Thorpe (2001) and Van Rullen and Thorpe (2002) presented a model for rapid face detection, where each neuron contributes only one spike. The authors argue that a spike based latency code is best to convey the required information from one processing stage to the next. Unfortunately, these models are strictly feed-forward and do not explain how processing continues beyond the first spike and beyond the first hypothesis.

There is, however, convincing experimental evidence that recognition in the visual cortex of mammals is a bidirectional, often top-down driven process (Bullier, 2001). Bidirectional models of sensory processing try to iteratively reduce the number of possible interpretations, using top-down prediction. A plausible mechanism of how prediction could improve the interpretation of a scene is the removal of already recognized parts from the input. Interpretation may then proceed with the residual parts (Barlow, 1994; Mumford, 1994).

This strategy is called *predictive coding* and is implemented in some models of visual processing (Koerner, Tsujino, & Masutani, 1997; Rao & Ballard, 1999). Indeed, these models could explain a number of experimentally observed phenomena, like various extra-classical receptive field effects (Rao & Ballard, 1999). But since predictive coding requires an initial hypothesis (prediction) to be available, it fails to explain how an initial stimulus hypothesis can be rapidly established.

In our view, the first hypothesis must be generated using only the fast forward connections and before feedback from higher levels arrives at the lower processing levels. Iterative refinement or predictive coding may then support discrimination within a recognized object class. We therefore suggest that the cortex uses two strategies to analyze sensory inputs: Faced with a new stimulus, the cortex first tries to catch the gist of the scene as quickly as possible. Only local decisions, that are with a high probability correct, are transmitted in a rapid feed-forward system. The gist is then fed back as a *global hypothesis*. A delayed feed-forward system transmits more ambiguous local decisions only if this feedback provides the bias towards a proper local decision within the context of the global hypothesis (Koerner, Gewaltig, Koerner, Richter, & Rodemann, 1999).

Since function and structure are not independent, the cortical architecture should reflect these two processing modes. There must also be mechanisms that determine which of the two modes is used at a given point in time. We propose that the clearly visible cortical layers implement the different processing modes.

The fast feed-forward system, which generates a first coarse hypothesis, may be in layers IV and lower layer III, and the slower refinement system in layers II/III (Koerner et al., 1999). The connections between layers bring these steps into the right order. In our view, a cortical column is, thus, primarily defined by the specific connectivity that links neurons in different layers into a functional circuit. This differs from the traditional view that merely see columns as a *vertical array of neurons* (e.g. Casanova, 2005).

Based on this reasoning, we present a model of cortical processing (*Cortext*) that assigns specific computational roles to the layers and columns in each area. *Cortext* extends an earlier proposal (Koerner et al., 1999) that did not provide an implementation as well as a recent model (Kupper, Knoblauch, Gewaltig, Koerner, & Koerner, 2007) which was based on rate-coded neurons, rather than on spiking neurons.

Although we use the visual cortex as an example, we try to capture the essential elements of sensory processing in neocortex that apply to all sensory modalities. With our model we want to address the following questions:

1. Given a new stimulus, how is a fast hypothesis established along the forward connections in a representation hierarchy?
2. How does the established global initial hypothesis help to support further processing of the stimulus?
3. Considering the very short duration of neuronal spikes, how can bottom-up and top-down signals, each encoded in a wave of spikes, be integrated without missing each other?
4. How reliable is recognition of known stimuli and how does the global hypothesis improve recognition of noisy or incomplete stimuli?

In the next section, we will describe our model of cortical processing in detail. We will then present simulation results, showing how a stimulus quickly triggers a first hypothesis on the stimulus. This hypothesis is then fed back to the previous processing stages to confirm or reject the local decisions. Unknown or noisy stimuli will cause specific activity which can be used to learn new stimuli, based on the previously stored information (not shown). Finally, we summarize the main results of our study and discuss them in the context of other models of cortical processing.

2. Model

In this section, we describe our model *Cortext* in detail. We start with the main assumptions on which *Cortext* is founded. We then describe the model parts, connections, and stimuli. The details of *Cortext* are summarized in Tables 1–6. The arguments supporting our model are given in the discussion and in Table 4. This separation of model description and model justification follows the recommendations of Nordlie, Gewaltig and Plesser (2009) for a good model description practice.

2.1. Overview

We assume that the areas along a sensory pathway (e.g. ventral path) implement a representation hierarchy. The first area represents a stimulus in terms of basic features, e.g. lines. The next area transforms this representation into a new one, based on common combinations of low-level features. This continues in each area, until the stimulus is represented as an abstract symbol at the highest area.

Each area is an array of identical feature detecting circuits called *macrocolumns* with adjacent receptive fields. Macrocolumns are composed of columns, each tuned to one of the features that could occur in the macrocolumn's receptive field. Columns respond to stimuli in their receptive field by generating spikes. The closer the stimulus is to the preferred feature of the column, the sooner it will respond. Thus, a short *spike latency* signals that the feature was detected with high confidence (Van Rullen & Thorpe, 2002). Columns that respond fastest also carry the most reliable signal. *Forward inhibition* suppresses the less confident columns that respond later. Thus, only the most reliably detected features reach the next area.

Table 1
Components of *Cortext*.

A: Model summary	
Populations	Three areas, V1, V2, IT, and retina
Topology	1D retina
Connectivity	Feed-Forward, feedback, lateral, recurrent
Neuron model	Leaky integrate-and-fire, fixed voltage threshold, fixed absolute refractory time (voltage clamp)
Synapse model	α-currents
Measurements	Spike activity, parameters of spike volleys

B: Populations		
Name	Elements	Size/Number
Area	Macrocolumn	3 (V1,V2,IT)
Macrocolumn	Minicolumn	9 (retinal positions)
Minicolumn	Sub-system	m, number of local feature
Sub-system	Neuron population E, I	3 (A1, A2, B)
E	Iaf Neuron	40
I	Iaf Neuron	10
N_{ext}	Poisson generator	for each neuron (184 050)
TH1	Spike generator	1
TH2	Poisson generator	for each neuron in V1-A1 (10 350)

Table 2
Connections of *Cortext*, part 1. Divergent patterns, symbolized by $1 \rightarrow N$, denote that each neuron from a source sub-system projects to N neurons from the target sub-system. 'Random" means that the N target neurons are chosen randomly from the target sub-system.

C: Connectivity 1 — overview					
Name	Source	Target	Pattern	Figure	Remark
NOISE	N_{ext}	All neurons			
STIM1	TH1	A1-V1		1(a)	
STIM2	TH1	A1-V1		1(a)	
FF1	A1	A2	Divergent, $1 \rightarrow 40$	1(a), (b); 2(a)	For all excitatory pools (E) for areas V1, V2, IT
FF2	A1	B	Divergent, $1 \rightarrow 40$	1(a), (b); 2(a)	For all excitatory pools (E) for areas V1, V2, IT
	V1-A2	V2-A1			
FF3	V2-A2	IT-A1	Random divergent, $1 \rightarrow 12$	1(a); 3	
	V1-B	V2-B			
FF4	V2-B	IT-B	Random divergent, $1 \rightarrow 12$	1(a); 3	
	IT-B	V2-B			
FB1	V2-B	V1-B	Divergent, $1 \rightarrow 40$	1(a); 3(e)	
FB2	B	A2	Divergent, $1 \rightarrow 40$	1(a); 3(b)	For areas V1, V2, IT
RC1	E	E	Random divergent, $1 \rightarrow 4$	1(d); 3(d)	For all sub-systems
RC2	E	I	Random divergent, $1 \rightarrow 1$	1(d); 3(d)	For all sub-systems
RC3	I	E	Random divergent, $1 \rightarrow 4$	1(d); 3(d)	For all sub-systems
RC4	I	I	Random divergent, $1 \rightarrow 1$	1(d); 3(d)	For all sub-systems
LAT			Random divergent, $1 \rightarrow 1$	3(c)	For all sub-systems

Table 3
Connections of *Cortext*, part 2. Synaptic weights in mV denote the amplitude of the post-synaptic potential (see Appendix).

D: Connectivity 2 — parameters			
Name	Weight	Delay	Remark
NOISE	0.1 mV	1.0 ms	6670 Hz Poisson process
STIM1	0.35 mV	1.0 ms	Three synchronous spikes
STIM2	0.35 mV	1.0 ms	300 Hz
FF1	0.35 mV	1.0 ms	
FF2	0.01 mV	3.0 ms	Up-modulated, maximal weight 0.35 mV
FF3	0.35 mV	3.0 ms	
FF4	0.01 mV	3.0 ms	Up-modulated, maximal weight 0.35 mV
FB1	2.0 mV	5.0 ms	
FB2	−2.0 mV	1.0 ms	
RC1	0.1 mV	0.5 ms	
RC2	0.1 mV	0.5 ms	
RC3	−0.2 mV	0.5 ms	
RC4	−0.2 mV	0.5 ms	
LAT	−0.2 mV	0.5 ms	

Once a global initial hypothesis is established in the highest area, it is fed back to the next lower area to *switch-off* all features that are part of the global hypothesis, such that the only more abstract description remains. This continues top-down until the stimulus is only represented at the highest level. Any residual activity in lower-level areas, thus, represents parts that could not be described by higher-level areas. Since the fastest and most reliable features are now suppressed, less reliable decisions may reach the next level to refine or correct the global hypothesis. We refer to this process as *switching-off*. It is consistent with the idea of *predictive coding*, discussed above.

Cortical columns host and coordinate the two parallel processing systems. Each of the two systems is embodied in distinct cortex layers:

1. A fast feed-forward system quickly generates a coarse, but reliable hypothesis, based on the fastest (and, thus, most reliable features). We call this the *A-System*. It is embodied by layers IV and III.
2. A delayed *feedback-mediated feed-forward system* iteratively refines the coarse hypothesis by resolving local ambiguities, using the global hypothesis, and by adding detail, using less reliable decisions. We call this the *B-System*. It is embodied in layers II/III.

2.2. Model composition

Cortext has a retina and three cortical areas, *V1*, *V2* and *IT*, each representing a step in visual processing. The general architecture of *Cortext* is shown in Fig. 1(a).

Table 4
Connections of *Cortext*, part 3.

E: Connectivity 3 — functional interpretation			
Name	Anatomical counterpart	Functional interpretation	Reference
STIM1	Thalamic input	Phasic stimulus part	(Blasdel & Lund, 1983; Steriade et al., 1997; Wang et al., 2006)
STIM2	Thalamic input	Tonic stimulus part	(Blasdel & Lund, 1983; Steriade et al., 1997; Wang et al., 2006)
NOISE	Embedding into network	Membrane potential below threshold	(Calvin & Stevens, 1968)
FF1	IV → III	Fast feed-forward (A)	(Callaway & Wiser, 1996; Fitzpatrick et al., 1985)
FF2	IV → II/III	Up-modulated feed-forward (A)	(Callaway & Wiser, 1996; Fitzpatrick et al., 1985)
FF3	II → IV of higher area	Feed-forward between areas	(Hirsch & Martinez, 2006; Nassi & Callaway, 2007; Van Essen et al., 1986; Wang & Burkhalter, 2007)
FF4	II/III → II/III of higher area	Modulated feed-forward	(Lorente de No, 1949; MacLeod & Laurent, 1996; Sawatari & Callaway, 2000)
FB1	II/III → II/III of lower area	Modulatory feedback	(Eckhorn et al., 1990; Kiebel et al., 2008; Rockland & Van Hoesen, 1994; Rockland & Virga, 1989; Sherman, 2007; Shmuel et al., 2005)
FB2	II/III → III of same area	Inhibitory feedback	
RC1-RC4	Recurrent connectivity		
LAT	Lateral inhibition	Winner-take-all inhibition	(Markram et al., 2004)

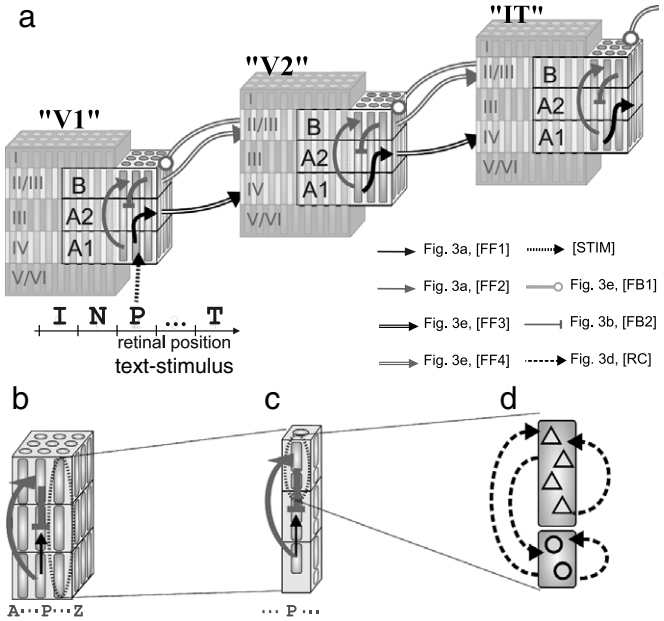


Fig. 1. The *Cortext* model. (a) Shown from left to right are the different areas of the *Cortext* model, V1, V2 and IT, together with an overview of their connections. Double lines depict connections between areas. They define the receptive field profiles of the target columns. Arrow-heads indicate excitatory connections, and round-heads depict modulating connections. They define two pathways: A fast forward pathway connects the three areas by their respective A- (black arrows) and B-systems (gray arrows). This defines the forward processing system. The second pathway connects the B-systems in a recurrent loop (gray round-heads). This defines the modulatory feedback system. (b) An area is divided into macrocolumns, each receiving stimuli from a single retinal position. Arrow-heads indicate excitatory connections, flat-heads indicate inhibitory connections and round-heads depict modulating connections. (c) Macrocolumns are again grouped into minicolumns. Within each minicolumn, the three sub-systems, A1, A2, and B interact via excitatory feed-forward (A1 → A2 and A1 → B) and inhibitory feedback (B ⊣ A2) projections. (d) Sub-systems are each implemented by a recurrent network of 40 excitatory and 10 inhibitory neurons, denoted by triangles and circles, respectively. The arrows denoting the connectivity are again shown together with references to Fig. 3 and abbreviations used in Tables 2–4.

Feature hierarchy and stimuli

The feature hierarchy of *Cortext* consists of letters, syllables, and words. These serve as a simplified metaphor for the much more higher-dimensional and also unknown, features of the visual system (see Discussion). Each level in this feature hierarchy is represented by one area. The first area V1, can recognize an alphabet of 23 letters. The second area V2 can distinguish 80 syllables with three letters each. The third area IT represents 300 words, composed of three syllables or nine letters. The task of *Cortext* is to recognize words from the sequence of letters presented to the retina. The complete set of letters, syllables and words is shown in the Appendix.

Table 5
Neuron model and data analysis.

F: Neuron and synapse model	
Name	Iaf neuron
Type	Leaky integrate-and-fire, α-current input
Subthreshold dynamics	$\tau_m \frac{dV(t)}{dt} - V(t) + V_0 + R_m \cdot I_{syn}(t)$ if $t > t^* + \tau_{rp}$ $V(t) = V_0$ else $I_{syn}(t) = \sum_i \sum_{s \in S_i} J_i \cdot \alpha(t - s - \delta_i)$ $\alpha(t) = t \cdot e/\tau_{syn} \exp(-t/\tau_{syn}) \cdot \mathcal{H}(t)$
Spiking	If $V_\uparrow(t) \geq \theta$ (threshold crossing from below) 1. Set $t^* = t$ 2. Emit spike with time-stamp t^*
Modulation	$J(t) = J_0 \times (1 + \mu(t))$, $\mu(t) = f_{mod} \sum_i \sum_{s \in S_i} \exp \frac{(-t-s-\delta)}{\tau_{mod}}$,
Parameters	$V_0 = 0.0$ mV, $V_\theta = 20.0$ mV, $\tau_m = 20$ ms, $t_{ref} = 2.0$ ms, $\tau_{syn} = 0.5$ ms, $\tau_{mod} = 50.0$ ms, $f_{mod} = 2.0$
G: Measurements	
Spike activity as raster plots	
Parametrization of spike volleys: number of spikes (a), time (mean, μ), temporal spread (standard deviation of spikes, σ)	

Table 6
Stimuli in *Cortext*. Letter sub-systems in V1-A1 are stimulated according to their overlap with the stimulus letter. An example where a single retina position is stimulated with the letter A (line 1) is shown. The overlap between the preferred stimulus (second row, gray) and the stimulus is illustrated in the third row and the common pixels in the forth row. For every preferred stimulus (line 2), similarity is expressed by the pixel-correlation c_i between stimulus and preferred stimulus (Eq. (9)). The synaptic weights are calculated by potentiating the correlations, $p_i = c_i^3$ (fourth row) followed by a normalization such that the sum of all weights is a constant. To this end, we multiply p_i by a factor $C = w_0 / \sum p_i$ such that $\sum w_i =$ const.

H: Illustration of the stimuli						
Stimulus	A	A	A	...	A	A
Preferred stimulus (PS)	A	B	C	...	Y	Z
Pixel overlap with PS	A	B	C	...	Y	Z
Common pixels	A	B	C	...	Y	Z
pixel covariance, $c_i =$	1.0	0.55	0.23	...	0.005	0.017
unnormalized weight $p_i = c_i^3 =$	1.0	0.16	0.01	...	0.00	0.00
normalized weight (in mV) $w_i = p_i * C =$	**3.8**	**0.37**	**0.30**	...	**0.00**	**0.0**

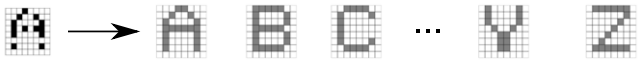


Fig. 2. Features in columns of V1. The receptive field templates for minicolumns in V1 are binary bitmaps of 8 × 8 pixels, forming 23 letters (gray). A stimulus occurs by determining the pixel-wise overlap between the input image (black) and the respective template. Features in V2 are syllables, 3 × 1 arrays of letters. Features in IT are 3 × 1 arrays of syllables, defining words.

Retina

The retina is modeled as a grid of 72 × 8 receptors, divided into 9 non-overlapping positions of 8 × 8 receptors. Letters are presented as 8 × 8 bitmaps to one of the positions on the retina. This is illustrated in Fig. 2.

Areas

The retina projects to the first area V1. From there activity can propagate via a fast forward pathway (double-lined arrows in Fig. 1) over area V2 to the highest area IT. Activity can be fed back to lower areas via the connections depicted with round-endings in Fig. 1.

All the areas have the same basic structure. They are one-dimensional arrays of *macrocolumns* (Fig. 1(b)) with a receptive field in the respective lower-level area. Neighboring macrocolumns have neighboring receptive fields.

Macrocolumns in V1 observe one position on the retina. Receptive fields in V2 cover three adjacent V1 macrocolumns, corresponding to three adjacent retina positions. Finally, the receptive fields in IT cover three positions in V2 or nine positions on the retina. Thus, from V1, over V2 to IT, the receptive field size on the retina increases.

Columns and sub-systems

Each column is selective to one of the features of its level (Fig. 1(c)). Macrocolumns group all columns that have the same receptive field and contain one column for each feature. For example, a macrocolumn in V2 contains 80 columns, one for each syllable. Columns within a macrocolumn compete, such that only the fastest responding columns can send their output to the next level.

Within a column, there are three sub-systems (Fig. 1(d)) that we call A1, A2, and B. Input from the previous level enters in A1 and is further propagated via feed-forward connections to A2 (black arrows in Fig. 1(a), (b)). A2 is the primary output system of a column and is connected to the A1-system of columns in the next area. The B-system implements feedback between and within columns as well as feed-forward connections to the next higher area (gray lines in Fig. 1(a)).

Each sub-system is modeled as a sparsely connected network of 40 excitatory and 10 inhibitory neurons as illustrated in Fig. 1(d).

Neurons

Neurons are modeled as integrate-and-fire neurons (Tuckwell, 1988) whose membrane potential is described by the following differential equation:

$$\tau_m \frac{dV(t)}{dt} = -V(t) + V_0 + R_m \cdot I_{syn}(t) \quad (1)$$

where τ_m denotes the membrane time constant, R_m the membrane resistance, and I_{syn} the total synaptic current. The membrane capacitance C_m is equal to τ_m/R_m. When V_m reaches a fixed threshold V_θ, a spike is emitted and the membrane potential is reset to the resting potential V_0 for the time of the refractory period t_{ref}.

Cortext has three types of synaptic connections between and within populations: excitatory, inhibitory, and modulating.

Only spikes at excitatory and inhibitory synapses affect the neuron's membrane potential, by producing excitatory (EPSC) or inhibitory (IPSC) post-synaptic currents.

Post-synaptic currents (PSCs) are modeled as α-functions (Jack, Redman, & Wong, 1981)

$$\alpha(t) = t \cdot e/\tau_{syn} \exp(-t/\tau_{syn}) \cdot \mathcal{H}(t) \quad (2)$$

where $\mathcal{H}(t)$ is the Heaviside function.

We then write the total synaptic current $I_{syn}(t)$ as:

$$I_{syn}(t) = \sum_i \sum_{s \in S_i} J_i \cdot \alpha(t - s - \delta_i), \quad (3)$$

where i runs over all pre-synaptic neurons and s over all spike times S_i of a pre-synaptic neuron i. The amplitude and the synaptic delay of the connection to the pre-synaptic neuron i are denoted as J_i and δ_i, respectively.

Modulating synapses enhance the excitatory post-synaptic potentials from other pre-synaptic neurons. This is done by scaling the synaptic weight amplitude J by a factor that exponentially decays to one (Eckhorn, Reitboeck, Arndt, & Dicke, 1990):

$$J(t) = J_0 \cdot (1 + \mu(t)), \quad (4)$$

where

$$\mu(t) = f_{mod} \sum_i \sum_{s \in S_i} \exp \frac{(-t - s - \delta)}{\tau_{mod}}, \quad (5)$$

and S_i are the spike times of pre-synaptic neurons that are connected with modulatory synapses (round-headed arrows in Figs. 1 and 3(e)).

The weights of modulated excitatory synapses are chosen such that an EPSP cannot bring the neuron threshold without an additional modulating input. In Fig. 1, modulated excitatory connections are denoted by gray arrows.

Synaptic weights and delays are different for connections within and between columns and areas. For each synaptic weight, we choose J such that the amplitude of the post-synaptic potential (PSP) has a desired value in mV (see Appendix).

To avoid an unstable feedback loop, modulated synaptic weights are clipped such that they cannot exceed the weights of the static excitatory connections. Tables 3 and 5 summarize the values used in our model.

Surrounding network

We assume that all neurons receive input from the surrounding cortical network. The rate and the synaptic weight of this background activity are chosen such that the membrane potential of all neurons fluctuates at a value 5 mV below firing threshold. The required rate of the Poisson process can be determined, using Campbell's Theorem (Papoulis, 1991) which states that the mean membrane potential μ, caused by a shot-noise process with rate λ_{noise} is given by:

$$\mu = \lambda_{noise} \int_0^\infty \mathrm{PSP}(t) dt, \quad (6)$$

where $\mathrm{PSP}(t)$ is the time-course of the post-synaptic potential, obtained by convolving the post-synaptic current $\alpha(t)$ with the membrane's impulse response $1/C_m \exp(-t/\tau_m)$. Using the closed expression of the integral

$$\int_0^\infty \mathrm{PSP}(t) = J \tau_{syn} e \tau_m / C_m, \quad (7)$$

and substituting the parameters with values given in the Tables 3 and 5, we arrive at a population firing rate of $\lambda_{noise} = 6\,670$ Hz.

Stimulation

The strength of an input to a V1 column is defined by the amplitude of its post-synaptic potential, which is determined by the inner product between the column's receptive field profile

and the stimulated pixels in its receptive field. For example, if the letter E is presented, the column representing this letter will be stimulated maximally. Other columns, representing similar letters, in this case F and L, will receive a stimulus that is correspondingly weaker.

Since we know the receptive field profiles of all V1 columns and also the set of possible stimuli, we can pre-compute the all possible inputs to a V1 column. Thus, we do not need to model the retina explicitly.

The time course of the stimulus mimics the response of LGN relay cells to visual stimulation. It usually has two parts. First, a sharp transient peak at stimulus onset, followed by a weaker tonic part. We model the initial transient by a number of synchronous spikes. The tonic response is modeled by 250 ms of random Poisson spikes with a constant rate of 300 Hz. Note that this population rate comprises the spikes of all neurons in the column's receptive field.

2.3. Connectivity

We now describe the connections between the model components of *Cortext*. The general connectivity is shown in Figs. 1 and 3. The details, functional interpretation, as well as the biological justification of the connectivity are given in Tables 2–4.

Overview

There are two routes along which spikes can propagate from the retina to the highest area IT. First, via fast connections that connect the sub-systems A1 to A2 in one area to the A1-system of the next area, as illustrated by the black arrows in Fig. 1(a).

The second route leads from A1- to the B-system. From there, the connections lead from one B-system to the next, up to IT, depicted by the gray arrows in Fig. 1(a). The forward connections from one B-system to the next are too weak to trigger activity in the respective B-system. Rather, they depend on input from the modulatory connections that arrive from the B-system of the next level. These follow the same route, only top-down. Thus, modulating connections start at the highest-level IT and connect the B-systems down to V1.

At the highest area IT, all B-systems receive unspecific modulating input ($\mu = 2.0$). Thus, once excitatory input arrives in IT-A1, IT-B will become active as the first B-system in the network. IT-B will then enable lower B-systems to be activated.

Once activated, the inhibitory neurons of the B-system will inhibit the A2 neurons of the same column. These connections *switch-off* of those columns that have contributed to an active column at the next level.

Preferred stimuli and receptive field

The forward connections to the A1-system of a column define the receptive field as well as the stimulus preference of the entire column. The forward connections from A1 to A2 and B relay these properties to the other sub-systems in the column (Fig. 3).

The stimulus preference is implemented by excitatory connections from the A2-systems of several lower-level features to the A1-system of one higher-level column. This is illustrated in Fig. 3(e) with the example of the word ANYANY. Columns in V1 receive input from one retina position. Columns in V2 and IT receive input from 3 columns in the respective lower level.

From each macrocolumn, only the column whose feature is best activated should signal its decision to the next processing level (area). This requires competition between the columns of a macrocolumn. *Cortext* has two mechanisms to select the best matching column: forward inhibition and lateral inhibition.

Forward inhibition

Forward inhibition is implemented by connections between adjacent areas. These emanate from the same A2 neurons as above, but target inhibitory neurons (in A1) of those columns

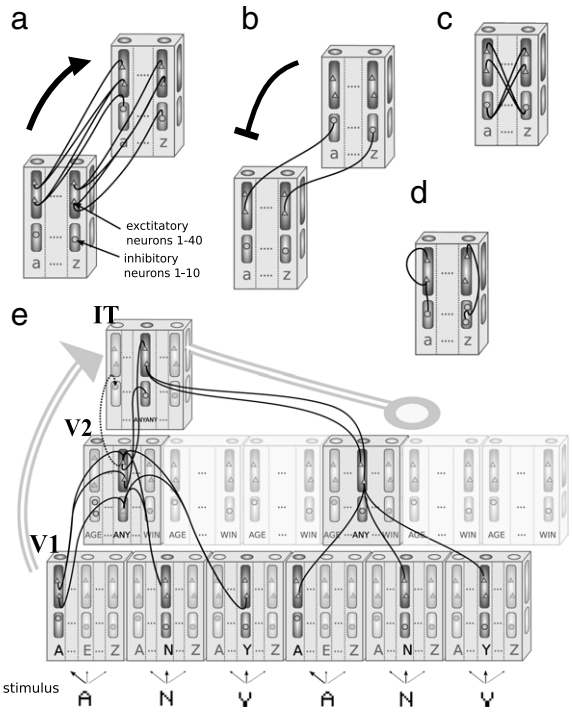


Fig. 3. Connections in the *Cortext* model. Arrows refer to connections symbolized in Fig. 1. (a) feed-forward connections. Each neuron from the excitatory source sub-system (triangles) projects to all neurons of the target sub-system (only few synapses are shown, for clarity). Both the source and target neurons represent the same feature (gray letters). (b) inhibitory feedback within one area. Interneurons (circles) project back to the excitatory target population. Again, synapses fully connect two sub-systems with the same feature representation. (c) lateral inhibition in all systems. Each representation inhibits all others by projecting inhibitory synapses from the sub-system's interneurons to five percent excitatory neurons from all other sub-systems (randomly chosen). (d) recurrent connectivity is implemented within and across excitatory and inhibitory neurons of each sub-system. The connection probability for each case is 10%. (e) explanation of the inter-area connections, denoted by double-lined arrows in Fig. 1. Sub-systems that represent features at a particular position on the retina are connected according to the word composition into syllables and letters. The connectivity for the word *anyany* is shown. The feed-forward and feedback connections are separately shown for the two syllables, for clarity, but apply to any syllable. Neurons from a sub-system of one level project to neurons in a target sub-system (black arrows). The probability that two neurons are connected is 0.3. Conversely, modulatory synapses project back to all neurons from the source sub-system (round-headed arrows). Connections are established for every word, however, multiple connections are excluded. Words were chosen to have a similarity (correlation coefficient) less than 0.85 to one another. In addition, each sub-system sends excitatory projections to the interneurons of other sub-systems (dotted arrow), whose features have a particular similarity to the feature represented by the target sub-system (see text). The retinal width in our model is nine features, shown is a width of six, for clarity. (a–d) shows examples for V1 where the local features are letters, but apply to each area.

that represent similar features. In *Cortext*, this forward inhibition is restricted to the ten percent most similar features. This is illustrated by the dashed arrow in Fig. 3(e).

Lateral inhibition

The columns within a macrocolumn inhibit each other through connections from the inhibitory neurons of all sub-systems (A1, A2, and B) in a column to the excitatory neurons of the corresponding sub-system in all other columns. Thus, the column which is activated first, will suppress all competing columns in its macrocolumn.

Modulating feedback

The feedback connections between the B-systems connect the same columns as the forward connections that define the preferred stimulus, but in opposite direction. Thus, they start at the higher-level column and target several lower-level columns.

Recurrent connections

Finally, neurons *within* sub-systems have recurrent connections (Fig. 3(d)), such that each neuron receives random connections from 10% of the excitatory and inhibitory neurons. In this random selection multiple occurrences are possible. (See Figs. 1(d) and 3(d)).

2.4. Methods

In total, *Cortext* consists of 27 macrocolumns (three areas with 9 retinal positions each). Depending on the number of local features, each macrocolumn comprises between 3000 and 45 000 neurons. Each of the 1227 minicolumns has 150 neurons. Thus, the entire network comprises about 190 000 neurons and 26 000 000 synaptic connections.

We simulated *Cortext* with the neural simulation tool NEST (Gewaltig & Diesmann, 2007), using its Python interface PyNEST (Eppler, Helias, Muller, Diesmann, & Gewaltig, 2008). The simulation data was recorded to disk and analyzed off-line using the Python libraries NumPy/SciPy (www.scipy.org).

Further details and the simulation parameters are given in Tables 1–6 and in the Appendix.

3. Results

In this section, we investigate how *Cortext* responds to sequences of letters, presented to the input system V1-A1. Fig. 4(a) sketches the time-course of the stimulus. It starts with a number of synchronous spikes at $T = 0$ ms, followed by a tonic phase of asynchronous input (spikes drawn from a Poissonian distribution) and lasts 250 ms. Fig. 4(b) shows the response of neurons in the input system V1-A1 to the stimulus HADFERSIM. A dot in the raster plot represents the time of spike of one neuron.

Each row in the plot shows the spikes of all neurons from the V1-A1 sub-system of a particular column, observing the first retinal position and with a preferred stimulus that is shown to the left of the row. For example, neurons in the first row respond maximally to the letter H, neurons in the second row prefer letter A, and so on.

All shown neurons increase their firing rate, corresponding to the degree of match between the stimulus and their receptive field profile. The stimulus perfectly matches the receptive field profile of the neurons shown in the top-row of Fig. 4(b) (black dots). These neurons respond strongest to the stimulus. The neurons in the other rows respond weaker (gray dots), since the stimulus only partially matches their preferred stimulus.

Fig. 4(c) shows the mean firing of neurons whose preferred stimulus is perfectly matched (black curve) and of the A1 neurons in other columns that responded with at least 10 spikes (gray curve). The optimally stimulated neurons fire around 41 Hz, while the sub-optimally stimulated neurons fire only at about 21 Hz. Moreover, the optimally stimulated sub-system responds to the initial transient of the stimulus with a sharp volley of spikes. We will see below that such spike volleys define the dynamics of the entire network.

We can characterize a spike volley by its time of occurrence, its number of spikes and its temporal precision, as illustrated in Fig. 4(d). The number of spikes measures the strength of the volley. In our analysis, we require that a volley contains more than ten spikes within a 5 ms time interval. Once we have identified a volley with this criterion, we collect all spikes within a 10 ms window around that time. The mean of all spike times in this 20 ms window defines the precise time of a volley and their standard deviation the temporal precision (Gewaltig, Diesmann, & Aertsen, 2001).

Fig. 4(f) shows the response volleys of all columns in V1 in the two-dimensional space, defined by their times of occurrence and their temporal spread. Black circles represent volleys in response to a preferred stimulus (black) while gray circles represent volleys in response to a non-preferred stimulus. We observe that the response to the preferred stimulus (*correct* response) occurs on average after 2.5 ms, while responses to non-preferred stimuli (*incorrect* responses) occur 2.9 ms later (at 5.4 ms). We also observe that volleys representing correct responses carry more spikes and are also more precise (in terms of their standard deviation).

Correct volleys contain on average 40 spikes with a standard deviation of 0.8 ms. Hence, all excitatory neurons respond almost at the same time. By contrast, the average incorrect volley contains only 29 spikes with a standard deviation of 2.2 ms.

In the following, we shall call a response *correct* if the first and strongest volley originates from a column whose receptive field profile matches the stimulus. Thus, the response shown in Fig. 4(b) is correct, because the neurons which represent the letter H also respond strongest to the stimulus H. Correct responses in V2 will originate from columns that represent syllables formed by the presented letter combinations. Finally, a correct response in IT will be from the column that represents the word which is formed by the stimulated letters.

If the stimulus is unambiguous, responses in V1 will always be correct, because of their direct input from the retina. In other areas, however, columns integrate activity from neighboring macrocolumns and the top-down signals.

3.1. Initial hypothesis

Fig. 5 shows the spiking activity in areas V1 (bottom), V2 (middle), and IT (top) during the first 25 ms. Each box represents a cluster of columnar sub-systems repeating the features at the respective retinal position (numbers on the left). The represented feature is also shown on the left of each row. Note that the rows correspond to the "retinal" positions indicated by the number.

First, we investigate whether the network is able to generate a fast hypothesis about the stimulus. To this end, we look at the response of those columns which actually represent the stimulus during the first 25 ms after stimulus onset (Fig. 4).

According to the inter- and intra-areal connections (Figs. 1 and 3), the signals generated in the input system of V1 (Fig. 4) propagate to subsequent sub-systems and areas. Fig. 5 indeed shows how a volley travels through all nine sub-systems (three per area). The following sub-systems respond with increasing delays. As in the input system of V1 (Fig. 4(b)) the other sub-systems respond with synchronous volleys of spikes.

Responses of columns that were initially dispersed in time are sharpened at each stage along the forward path, until in area IT, columns respond only with sharp volleys at around 20 ms. Columns that do not match the stimulated word (or its syllables and letters) also respond with volleys to the stimulus, albeit weaker (not shown in Fig. 5).

System B of areas V1 and V2 are almost silent. IT-B responds only with volleys, they are shown, along with incorrect responses, in Fig. 6.

Fig. 5 shows that in the column that represents the stimulated word (black), IT-B responds at around 21 ms with 40 spikes and a precision of 0.15 ms. This response is maximal, since all excitatory neurons participate in the volley.

In addition to the correct column, three other columns (gray) respond with volleys. This is shown in Fig. 6. We refer to the entire set of feature responses (correct and wrong) as initial candidate list which represents the first hypothesis of the network. In this example, the candidate list comprises 1.3% of the vocabulary of 300 words.

The volleys of the incorrect responses are also maximal (40 spikes) with a standard deviation only slightly larger than that of the correct response (0.19 ms). However, the correct response is fastest, occurring on average 1.5 ms earlier than the remaining responses.

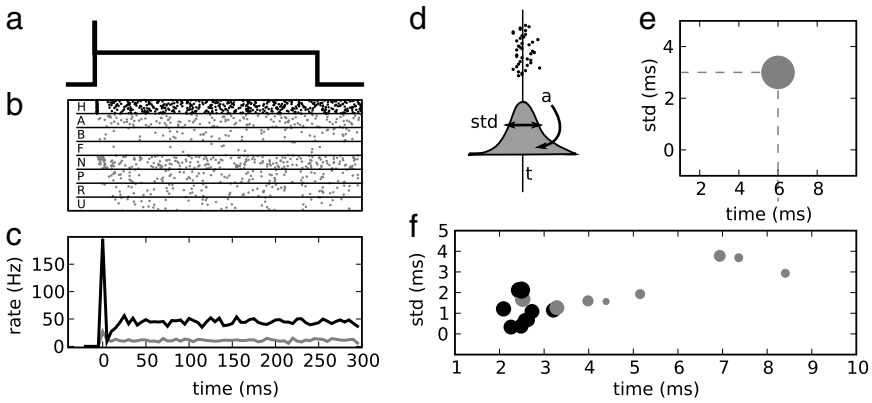


Fig. 4. Stimulus responses in V1-A1. (a) Time course of the stimulus. After applying three synchronous spikes at $t = 0$, the stimulus continues at a fixed rate for 250 ms. (b) raster plot of sub-system responses in V1-A1 for the first retinal position. The stimulus forms the word HADFERSIM. Each row (separated by horizontal lines) represents the excitatory V1-A1 neurons of a column observing the first retina position. Their preferred stimulus is indicated on the left. The preferred stimulus of the top row (black) corresponds to the stimulated letter, the remaining rows to the spike activity in columns with similar, but different preferred stimuli. The remaining columns did not respond significantly to the stimulus. (c) average population activity (in spikes/s) for the correct (black) and the incorrect (gray) letters. (d) characterization of the initial volley by determining the mean $t = \mu$, the standard deviation, std, of the spike times and the number of spikes, a, in a volley. (d) Each volley can be represented by a point in a coordinate system spanned by the occurrence time t (x-axis) and the precision std of a volley (y-axis). For each volley, we draw a circle with a diameter that corresponds to the number of spikes in the volley. (f) summary of all initial volleys of correct (black) and incorrect (gray) responses, if present. The volleys are obtained from all nine retinal locations. In each plot, the x-axis denotes the time (in ms) relative to stimulus onset.

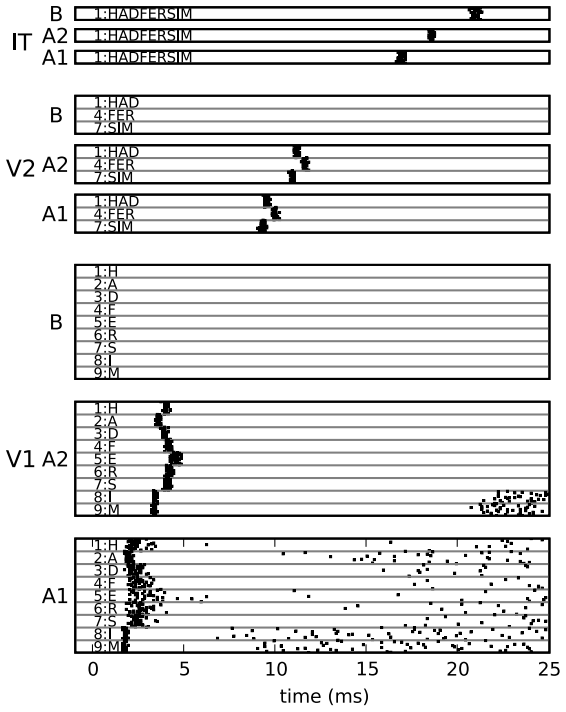


Fig. 5. Activity propagates through the systems. Each row (separated by gray lines) shows a raster plot of one sub-system representing the highlighted feature at the respective retinal position. Only responses of correct sub-systems are shown. The stimulus word is the same as in Fig. 4. Note that V1-A1 resumes firing after a pause (at 15 ms), while other responses consist of one volley only.

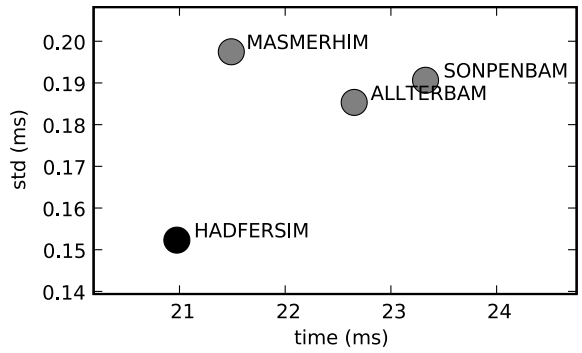


Fig. 6. List of initial candidates in IT-B. Dots depict the occurrence times (x-axis) and temporal precisions (y-axis) of volleys in sub-system B of area IT after stimulating the letters that form the artificial word HADFERSIM. Marked in black is the volley from the correct word-representing sub-system, while wrong volleys are plotted in gray. Time is shown relative to stimulus onset. The stimulus is the same as in Fig. 4.

To assess how the network generates its initial hypothesis on average, we ran simulations for 100 different and randomly chosen stimulus words. The responses are summarized in Fig. 7. In each case, the correct word and a small list of candidates is activated in IT-B, around 20 ms after stimulus onset. On average, the initial candidate list contains 4 words (1.33% of all learned words), always including the correct word. Correct sub-systems respond on average at 21 ms, while the wrong responses occur at 30 ms. Thus, the average correct response is ahead of their incorrect counterparts by 9 ms. The number of spikes in a volley is larger (40.0 spikes versus 38.2 spikes), and the widths of the correct responses are more narrow (0.18 ms standard deviation versus 0.30 ms). Thus, in all cases, the correct words responses are earlier, stronger, and more precise than their incorrect alternatives.

3.2. Refinement

We now consider the network response to the remaining 250 ms of stimulation. During that period, the stimulus consists of asynchronous Poisson spikes whose weight corresponds to the overlap between the stimulated letter and the neuron's receptive field profile.

The response to the initial stimulus transient is propagated only by the feed-forward connections between the A-systems of each area. After about 20 ms, this activity reaches the B-system of the highest area IT. All neurons in IT-B receive modulating input (round-headed arrow to IT-B in Fig. 1) and, thus, the forward activation is able to trigger activity in these neurons. Now, activity can propagate backwards along the feedback connections. We call this period *refinement phase*, the network response is shown in Fig. 8.

During refinement, volleys in IT-B are propagated back via modulatory synapses (round-headed arrows in Fig. 1) to V2-B. Thus, activation in IT-B that represents a word modulates the

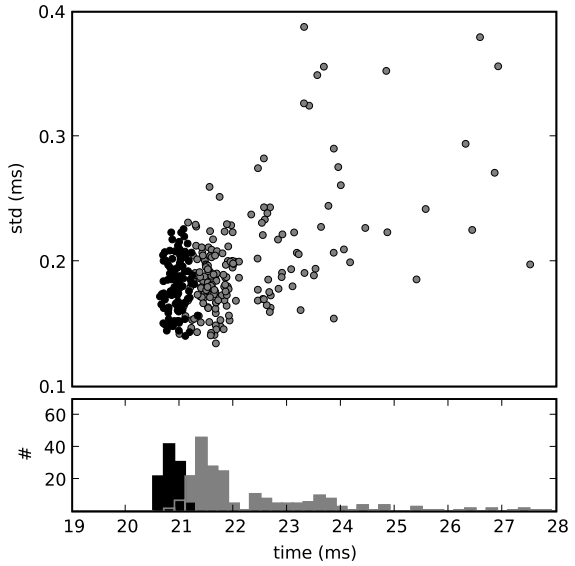


Fig. 7. Early feature activity in IT-B. The occurrence times (x-axis) and precisions (y-axis) of word-representing volleys for 100 recognition trials are shown in the upper panel. Each point denotes a different volley occurrence. The strength of the volleys (number of spikes) is denoted in the point diameter, the positions on the x-axis correspond to their occurrence time, and the position on the y-axis the temporal spreads. Correct words are emphasized in black, wrong words are gray. Time is shown relative to stimulus onset. The lower panel summarizes the occurrence times of the volleys in two histograms, correct (black) and wrong (gray) volleys. The bin width in the lower panel is 0.2 ms.

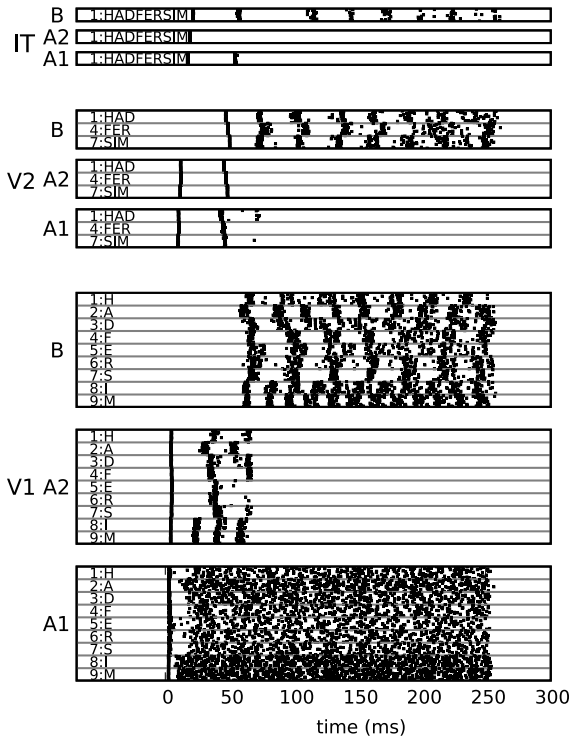


Fig. 8. Sub-System responses during refinement. As in Fig. 5, rows denote raster plots of sub-systems. Shown are the responses that are generated during the entire stimulus duration. The initial volley in IT-B enables responses first in V2-B, then in V1-B. For all areas, A2-activity stops at the onset of B-activity. Activity in the A1-system of areas V2 and IT stops after offset of the underlying A2-activity. V1-A1 remains active during the whole stimulus duration. Note that, apart from the network's input system (V1-A1), activations occur only in the B-systems.

synaptic weights to the three columns in V2-B that code for the syllables of the word. Responses occurring in V2-A1 can now reach the sub-systems in V2-B. They elicit spike volleys at around

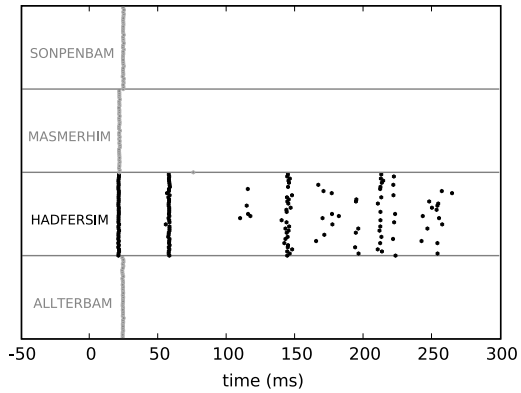


Fig. 9. Word representation. Raster plot of responses in four sub-systems in IT-B that respond with high activity. Marked in black is the correct activity (the sub-system representing the stimulated word), gray shows incorrect activity. After removal of the stimulus, the entire network becomes inactive. The network was simulated for 400 ms.

60 ms (Fig. 8) which are in turn propagated back to V1-B to modulate *its* afferents: columnar sub-systems in V1-A1. Finally, at around 70 ms, each columnar sub-system within the B-system that represents the feature at the retinal input has started to generate volleys.

Due to the strong feedback inhibition within each area ($B \dashv A2$), responses in the A2-systems are switched-off at the onset of B-activation. Therefore, volleys are no longer propagated via the A-A path, but rather via the modulated B-B connections. Apart from V1-A1, all other A-systems are silent for the rest of the stimulation, and volleys remain propagated between B-systems. Note that Fig. 8 contains only columns that represent the stimulus, wrong activations are not shown.

We now investigate responses in the word-representing system, IT-B. Fig. 9 shows raster plots of columns that respond most strongly. After the initial hypothesis, many of the neurons spike little or not at all. Only one column, the correct one (black), remains active, and its neurons fire synchronously. After the stimulus stops at 250 ms, the entire network becomes silent. Fig. 10(a) shows the responses of IT-B in terms of the volleys that were generated by the sub-systems. The initial wave of volleys at 20 ms is identical to Fig. 7. After the first wave at 20 ms, volley activity is resumed at about 60 ms after stimulus onset.

The occurrence time of the volleys is summarized in two histograms in Fig. 10(b), for the correct (black), and the incorrect (gray) responses. The volleys cluster around four times (22.3 ms, 59.7 ms, 117.8 ms and 147.1 ms, dashed lines in Fig. 10(a), (b)). Thus, during stimulation, area IT responds with waves of volleys that are 30–60 ms apart.

The second wave at 60 ms shows also volleys which we define as a second candidate list. While the first candidate list has many wrong responses, the second wave contains almost only the correct feature.

The amount of cases where the correct volley failed to reoccur is 3%, in 2% of the trials was IT-B activity uniquely but *wrongly* sustained.

Discernible peaks in the histogram of Fig. 10(b) enable us to define a reference time for volleys: As the two first waves do not overlap, we can use in both cases the *first* volley time as reference and measure subsequent volleys relative to this first volley. We do this, irrespective of whether the first volley is correct or wrong. Afterwards, the relative occurrence times can again be separated into correct and wrong responses. Fig. 10 shows the two histograms for the first (c) and the second (d) peak of the distribution in (b). The ordinate is the fraction of volleys from 100 trials. The origin ($t = 0$) corresponds to the time of the first volley.

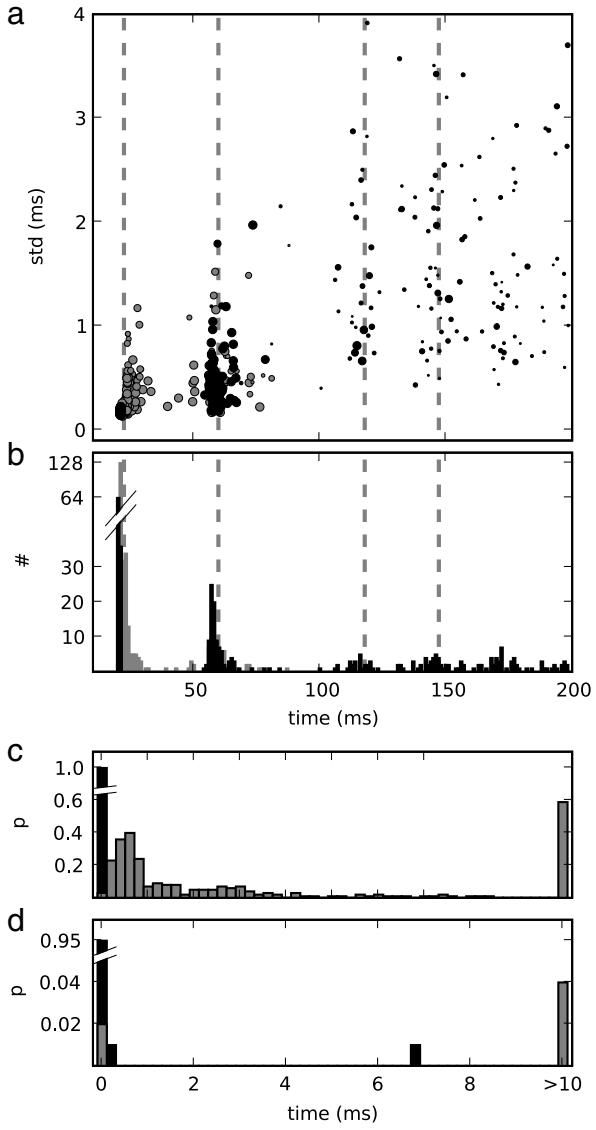


Fig. 10. Volley activity in IT-B. (a) volley symbolized by dots as in Fig. 7, showing the entire simulation duration. After the initial volleys at 20 ms, volleys occur in three further clusters (dotted lines). (b) histogram of correct (black) and wrong (gray) volley times. (c, d) histograms of correct (black) and wrong (gray) volley times, relative to the first volley occurrence. (c) shows the relative latencies of the initial cluster at 22 ms. A value of one for the correct (black) responses at $t = 0$ denotes that all sub-systems that responded at first were correct. Thus, all subsequent responses are false-positives (gray) (d) relative time histogram for the second cluster at around 60 ms.

$t > 0$ denotes volleys that occurred t milliseconds after the first volley. Therefore, if a wrong volley occurs at $t > 0$, it must have been preceded by a correct one at $t = 0$. Conversely, if a correct volley occurs at $t > 0$, it must have followed a wrong one at $t = 0$. By definition, the relative amount of correct and wrong volleys at $t = 0$ adds up to one.

For the first wave of volleys, a correct volley is the first to occur in all trials ($p = 1$, Fig. 10(c)). Consequently, all subsequent volleys are wrong responses (gray bars at $t > 0$).

The majority of the earliest volleys during the *second* wave is still correct ($p = 0.95$) (Fig. 10(d)), the remaining 5% did not occur or followed a wrong response. Only a small fraction (0.06%) of the first volleys is wrong, or occurs more than 7 ms later (1.33%).

At $t = 0$, Fig. 10(c), (d) define an *error rate*, given that only the first volley is considered. In the first case, the error is 0%, in the second case 5%. Moreover, the relative amount of false-positives

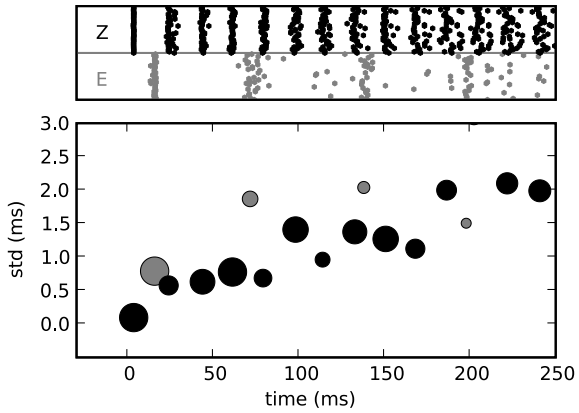


Fig. 11. Activity in V1-A2 for a "wrong" feature at position two. Shown in the upper panel are spikes that are generated in system A2 of area V1 after stimulation of the wrong word HZDFERSIM that contains an error at the second position as opposed to the known word HADFERSIM. The second retinal position shows strong responses for the sub-system representing the allegedly correct letter Z (black), together with a weaker activation of another sub-system, representing a similar letter, E (gray). The sub-system representing the actually correct letter, A, is not active. The lower panel shows the volleys as dots in the two-dimensional space. (Fig. 4(d), (e)).

(number of wrong versus correct volleys) is reduced in the second wave (from 34% to 16%).

3.3. Recognition stability and error signals

Due to the B ⊣ A2 inhibitory feedback (Fig. 3(b)), strong activity in the B-systems is able to suppress A2-activity. If the network is stimulated with complete and known stimuli, the activity in the B-systems is largely correct and inhibits the underlying A2-activity. If, however, one of the B-systems is not activated, e.g. because of an error in the stimulus, the according neurons in A2 will not be inhibited and therefore remain active.

Fig. 11 demonstrates that an error in the word stimulus indeed results in a strong activation of the respective V1-A2-sub-system at the corresponding position. In V1-B, however, this "wrong" sub-system is not activated, whereas here the correct sub-system is activated instead (letter A, not shown). Thus, incomplete stimuli can be completed by the predictive feedback from higher stages.

If a random word is presented to the network (an unknown word that cannot be segmented into known syllables), all B-systems remain silent whereas the according sub-systems in V1-A2 remain strongly activated during stimulation (not shown).

Realistic stimuli will only rarely be complete and correct. Thus, reliable recognition must also proceed in the presence of noise. To demonstrate the recognition stability of *Cortext*, we present incomplete stimuli. To this end, we randomly remove pixels from each stimulus letter with a particular probability p. This is done for fifty recognition trials.

Fig. 12 shows the relative amount of candidates for the first (dashed line) and the second (continuous line) wave of volleys versus the letter 'completeness' $1 - p$. The number of candidates is constant down to a completeness of 0.5 (i.e. 50% of the pixels are missing), then decreases for smaller values. During the refinement phase, the amount of wrong responses is decreased for all letter dilutions.

At a completeness of 0.5, the first pulse to occur is the correct one in 95% of the cases (for both waves, not shown).

3.4. Resolution of ambiguity

The previous results demonstrate how a set of candidate representations are activated, followed by a reduction to a single activation. However, it appeared up to now that the unique

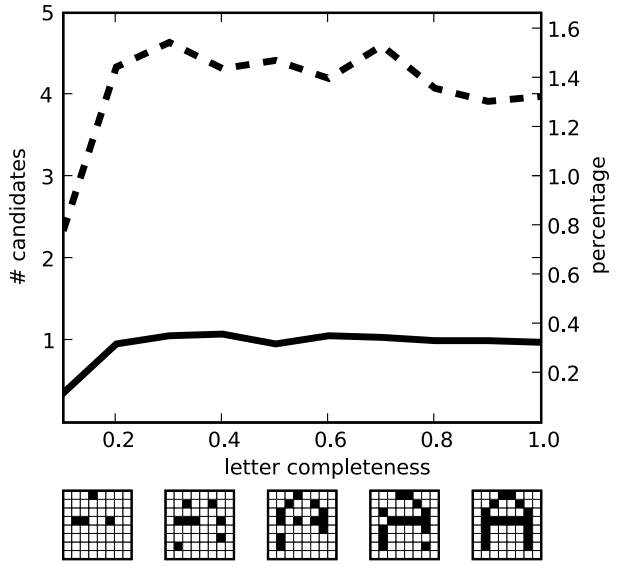


Fig. 12. Average amount of word-candidates in relation to random pixel dilution. All letters from the stimulus word were diluted by randomly removing pixels with a probability p. Shown on the x-axis is the letter 'completeness' $(1 - p)$. The letter bitmaps below are examples of stimuli, their dilution corresponds to their respective x-axis location. The y-axes indicate either the absolute number (left axis) or the relative amount of candidates (in percent, right axis), averaged from fifty simulation trials. The dashed line shows candidates that originated in the first wave of volleys, the continuous line refers to the candidates of the second wave.

activation of representations in IT-B is just predetermined by the order in which the initial hypothesis arrives: The earliest activation is the one that becomes sustained during the refinement phase. On the other hand, refinement of an initial hypothesis should also include error correction, going beyond a mere disambiguation of responses.

To test whether *Cortext* can correct a previously wrong hypothesis, we activated the columns in V1-A1 with noisy stimuli where a fraction of pixels was randomly chosen (60% of all pixels are chosen to be black or white with 50% probability). In order to allow a larger set of initial hypotheses, feedback inhibition (from B to A2) was reduced to −0.2 mV.

Simulations with 100 randomly chosen known words revealed that Cortext still generated a correct initial hypothesis in 87% of the trials. In 10% of the trials, the initial hypothesis was incorrect (activation of the correct word-representation did *not* occur with the smallest latency) but was later on corrected by sustaining correct activation (not shown). Thus, our model is indeed capable of correcting previously wrong stimulus hypotheses.

4. Discussion

In this contribution, we have presented *Cortext*, a columnar model that demonstrates the formation of a fast initial stimulus hypothesis, and its subsequent refinement by inter-columnar communication.

4.1. Recognition dynamics

The activation dynamics within the hierarchy of *Cortext* is summarized in Fig. 13. Panel (a) depicts the feed-forward path. Panels (b) through (d) illustrate how the global hypothesis, activated in IT-B propagates down and switches-off all A2-systems in lower levels which match the global hypothesis. Panel (d) shows the self-enforcing loop between the B-systems which stabilizes the percept at the highest level. Residual activity in any A2 hints at a mismatch between the global hypothesis and the local response.

Forward processing

Input from the retina enters the network in V1-A1. At stimulus onset, target sub-systems in V1-A1 respond with synchronous spikes (spike volleys) that are more or less pronounced, depending on the match between receptive field template and stimulus. A sub-system whose receptive field does hardly match the given stimulus shows only weak or no activity., i.e. no volleys are generated. In contrast, sub-systems that match the stimulus better respond with strong and precise volleys. The sub-system with the best match (whose receptive field is congruent with the stimulus) shows the strongest *and earliest* response (Fig. 4(b), (c)). Firing threshold is reached earlier due to the finite rise time τ_{syn} of the post-synaptic current Eq. (1), (Gewaltig et al., 2001).

Thereafter, the volleys propagate via strong connections to the next system (V1-A2) which in turn elicit spike volleys. At this stage, the activity in A2 is a delayed copy of the spikes in A1. The macrocolumns in V1 have generated a response which we call a local decision on the stimulus, or a *local hypothesis* — an activation that translates the match with local features in *strength and timing* of responses. The local features in V1 are letters, therefore its local decision contains information on which letters may have been stimulated.

Responses are propagated to the next area from V1-A2 (the output system of V1) and arrive at V2-A1. The sub-systems again respond according to the match between stimulus and *their* receptive field profile which here represent *syllables*. The match is now determined by the converging connections that originate from V1. These connections implement learned knowledge about how low-level features (letters) contribute to high-level features (syllables). Sub-systems in V2-A1 that code for a correct syllable receive more volleys from V1-A2 than other columns. In addition those volleys arrive also earlier. Columns that do not match at all remain silent. Hence, the responses in V2 represent local decisions on which syllables could be composed from the stimulated letters. Again, correct responses occur earlier than incorrect ones.

Local decisions of V2 are finally propagated to IT-A1, whose responses in turn code for the probability of words that may have been formed from the stimulus. IT-B is the first B-system to respond with spike volleys (Fig. 13(b)) because unlike the other B-systems, it receives unspecific modulation (Fig. 1(a)). This unspecific modulation mimics the effect of top-down attention (Itti & Koch, 2001) induced by the appearance of a new word.

At this stage, each area has made a decision on what local feature represents the stimulus best, in terms of letters (V1), syllables (V2), and words (IT). The local decision in IT stands for the few words that come into question, given the stimulated letters. This candidate list comprises about 1% of all learned words and represents a fast coarse estimate, sharing a high amount of overlap (cf. similar words in Fig. 9).

Forward inhibition and lateral inhibition

Fig. 13 does not show the two inhibitory mechanisms which sharpen the response properties of individual columns. Lateral inhibition acts between all columns of a macrocolumn and reduces the number of columns that respond to a stimulus to only the strongest (winner takes all). Forward inhibition reduces the input to those columns whose preferred stimulus is not perfectly matched. Experimental results provide evidence for both mechanisms (see e.g. (Ferster & Lindström, 1983; Porter, Johnson, & Agmon, 2001).

Feedback-mediated processing

Once activated, sub-systems in IT-B inhibit neurons in A2 and modulate the lower level's B-system (V2-B). Modulation amplifies all connections that enter V2-B: from V2-A1 and connections from V1-B. At first glance, this modulation seems to have no effect, as

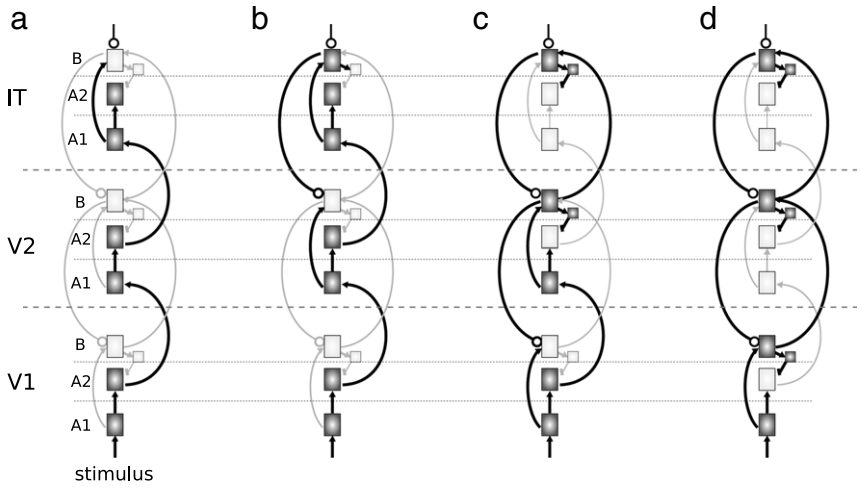


Fig. 13. Sequence of activations in the *Cortext* model. This schematic illustrates how sub-systems (denoted by boxes) are activated (emphasized in gray). To illustrate the general flow of activity, we neglect the feature specificity of each area and emphasize a box in gray, if the sub-system is activated in at least one column. Arrows are the connections (cf. Fig. 1), a connection which transmits activation is emphasized in black. (a) Initial activity propagates via the strong A–A connections up to IT-A2. Since the A1-B connections are weak, the B-systems at each level cannot be activated yet. (b) The B-system of IT receives unspecific modulation. This increases the A1-B connections sufficiently to trigger activity in IT-B, activating the interneurons in IT-B (small box) as well as modulatory feedback down to V2-B. (c) Switching-off starts in area IT, where the A2-systems are inhibited, followed by V2, where activity that has reached B de-activates V2-A2. (d) Final state after refinement. Apart from the input system, activity is only present in the B-systems.

V2-A1 has already responded with volleys. However, the stimulus remains active and neurons in V1-A1 resume firing after a short pause (Fig. 4(b)). Reoccurring volleys are now propagated to V2-B because their afferents have been selectively up-modulated by IT-B.

At this stage, volleys to IT are triggered directly by V2-B because the connections are modulated and activity in V2-A2 has been switched-off by the volleys in V2-B. Thus, the path along which responses are propagated has changed from one that is mediated by the A-systems, to one that is implemented by the B-systems. The same switch occurs in V2 with the first volleys in the sub-systems of V2-B at 60 ms and in V1 at 70 ms. Finally, volleys are propagated via the now strengthened connections between B-systems, and the primary feed-forward path becomes unused.

4.2. Stimulation without initial synchrony

It may be argued that the synchronization of responses in our *Cortext* model may depend on the initial transient of the stimulus. However, as Fig. 8 suggests, volleys can still occur during the second phase of the stimulus where the rate is constant. Responses are sharpened along areas, leading to more pronounced volleys.

Indeed, control experiments without initial synchrony showed a very similar network response, activating precise volleys after stimulus onset. Initial volleys occur on average 5.5 ms later if the stimulus consists only of Poissonian noise (not shown). The strength and precision of the initial volleys are, however, the same. We conclude therefore that an initial transient in the stimulus is not required to obtain a latency code. The sharp transient merely aids in shaping an earlier hypothesis.

4.3. Visual hierarchy and areas

Cortext foots on the assumption that the brain recognizes complex objects in a hierarchy of areas, each processing a specific class of stimulus features. This interpretation has a long history that dates back to the work of Sherrington (1941) and was eloquently summarized by Barlow's *Neuron Doctrine* (Barlow, 1972).

Work by Felleman and Van Essen (1991) showed that the visual system is indeed a highly interconnected graph of areas. Other authors have tried to extend the traditional view on visual processing by adding top-down driven processing (Hochstein & Ahissar, 2002) or by interpreting the visual system as a dynamic rather than a hard-wired hierarchy (Bullier, 2001).

Most models of visual processing build on the earlier ideas and interpret the visual system as essentially feed-forward (Delorme & Thorpe, 2001; Riesenhuber & Poggio, 1999).

4.4. Feature hierarchy

It has been suggested that systems processing visual information may use a *visual alphabet* along with a *visual grammar* according to which a stimulus is parsed (Watt & Phillips, 2000; Zhu & Mumford, 2006). Yet the details of such a feature hierarchy are still unknown. In our hierarchy, letters are the simplest features, next are syllables, formed by three letters, and finally, three syllables form words.

Many authors have tried to extract realistic features from the statistical properties of natural scenes (Olshausen, Anderson, & van Essen, 1993). However, this approach cannot be used beyond the first visual area.

By contrast, our approach starts from the essential properties of a visual feature hierarchy: (i) an infinite number of stimuli can be generated from a finite set of features, and (ii) features at a low level can be used to express the features at the next higher level.

Clearly, the visual areas V1, V2, and IT do not host the representations of written or spoken words. Rather, we want to illustrate principles of cortical processing by means of this well-defined, generative feature hierarchy.

4.5. Columns

Mountcastle (1997) was the first to propose columns as the general organization principle of the cortex. Since then, the idea has been under constant debate (Horton & Adams, 2005; Rockland & Ichinohe, 2004).

Recent work by Rakic (2008), however, emphasized both aspects that we consider important for columnar circuits: Columns not only vertically connect the cortical layers, but also share their extrinsic connectivity. They may thus be the basic functional units not only of sensory, but also of motor, and association areas.

4.6. Relation to other models

Riesenhuber and Poggio (1999) summarized much of what is known about hierarchical models of object recognition. They, and others (e.g. Wersing & Körner, 2003), demonstrated how objects can be recognized in a converging hierarchy of areas, each detecting and then pooling more and more complex features. Delorme and Thorpe (2001) and Van Rullen and Thorpe (2002) presented a model for rapid face detection, where neurons on each level contributes only one spike. The authors argued that a spike based latency code is best to convey the required information from one processing stage to the next (Thorpe, Delorme, & Van Rullen, 2001). Our *Cortext* model extends these models by introducing a feedback driven phase in which a rapidly generated initial hypothesis can be refined or even corrected (Section 3.4). *Cortext* also uses spike-latencies to encode information. For neurons, this code is easier to interpret than the *Rank Order Code* proposed by (Delorme & Thorpe, 2001; Van Rullen & Thorpe, 2002).

Grossberg and Versace (2008) recently proposed a spike based extension of the older LAMINART model of learning and attention (Grossberg & Raizada, 2000; Raizada & Grossberg, 2003), called SMART. They demonstrated how neurons can enter a resonant state, grouping low-level features in terms of spike coherence. There are two differences that distinguish our *Cortext* model from SMART. First, our Cortext model generates the first hypothesis in one rapid feed-forward sweep, using one spike per neuron, thus, no resonant state is necessary to establish an hypothesis at the highest level. Second, the resonant state of SMART emphasizes low-level features that are confirmed by the higher levels. By contrast, our *Cortext* model emphasizes the difference between two adjacent levels. The residual activity will then become important in two ways: first, it is a signal that a stimulus could not be completely explained. Second, it can be the trigger and basis for learning new representations, based on available representations and their difference to the stimulus (the residuals).

4.7. Relation to predictive coding

The mechanism of *switching-off* is very similar to predictive coding, as proposed by Rao and Ballard (1999), with one difference. In Rao's model, the difference between prediction and stimulus is passed from one processing level to the next. For the visual cortex this would mean that V2 would have to interpret an error signal generated by V1. This is not plausible for two reasons: First, receptive fields in V2 are larger than in V1. Second, in a feature hierarchy, a higher area would not be able to represent the error signal with its own abstract features. For example, in our Cortex model, residuals in V1 will be arbitrary combinations of letters. Since only some letters make up syllables of the next level, V2 will not be able to *see* the error signal of V1. By contrast, in the *Cortext* model, the error signals remain in the area where they naturally occur. As mentioned above, they can then be used in several ways, e.g. to raise attention or to drive learning.

4.8. Time code and reliability

Adrian (1928) was among the first to demonstrate that the discharge rate of a neuron corresponds strength of a sensory stimulus. In particular, for the early visual system it was assumed that information is encoded in the firing rate of neurons (e.g. Hubel & Wiesel, 1968). However, Thorpe et al. (1996) pointed out that the rate-coding hypothesis is not consistent with the high speed of visual processing.

Latency code, as we propose it, requires that neurons can produce action potentials with high temporal precision. Recent evidence by Butts et al. (2007) indeed suggests that in the visual system the timing of action potentials can be precise down to the millisecond. This leaves the question whether other timing differences could be detrimental to the latency code proposed here. One source of such differences could be the *off*-response of neurons, e.g. at the end of a fixation period. However, the off-response would occur 250–300 ms after stimulus onset, when the initial hypothesis has already been established. The off-response would also mark the end of a fixation and the beginning of a new saccade. Thus, the effects of the off-response will probably be masked by saccadic suppression. A system tuned to temporal signals may also use the temporally coherent off-response to refresh the temporal reference of the latency code.

The delayed responses of *lagged* cells (Saul, 2008), observed in the LGN may also be detrimental to a code based on spike latencies. However, these cells seem to be primarily involved in motion processing, which is not the domain of the model presented here.

Our model also links the response latency of a neuron to the reliability of its firing. This is supported by experimental evidence showing that neurons are most reliable and have the shortest response latency if they are stimulated optimally, both in the sensory periphery (Johansson & Birznieks, 2004), and in sensory areas (Heil, 1997; Osborne, Bialek, & Lisberger, 2004).

4.9. Experimental predictions

Our hypotheses regarding the roles of different cortical layers suggest a vertical heterogeneity of activity during recognition. Extracellular radial recordings of awake behaving animals through layers II–IV in areas V1/V2 may be able to find a differentiated stimulus-evoked response. However, there are very few experimental papers that distinguish at all between recordings in layers II and III (Gur & Snodderly, 2008). To verify our predictions, the animal should not see well-known moving bars but complex objects that need several levels of hierarchy for recognition.

Moreover, our A2-system suggests that activity in layer III should *decrease* following correct object recognition. A recent fMRI study found indeed elevated activity in the lateral occipital complex (LOC), accompanied with a *decrease* of activity in V1 during object recognition (Murray, Kersten, Olshausen, Schrater, & Woods, 2002). Later studies also observed reduced activity if predictions from higher areas matched the stimulus (Furl, van Rijsbergen, Treves, Friston, & Dolan, 2007; Harrison, Stephan, Rees, & Friston, 2007). Unfortunately, the measurements did not have a vertical resolution.

Selective inactivation of cortical layers (e.g. by careful cooling) may contribute to our knowledge about the function of layer II/III (B-system). Inactivation of layers II/III in IT should leave intact the generation of an initial hypothesis (measurable by reaction times), however impair more detailed recognition of noisy objects. In a comparable work, Lomber and Payne (2000) demonstrated the different functions of superficial and infragranular layers in motion-selective cortex of cats. He could indeed show that detection performance was left intact when only upper layers were disabled.

As each neuron contributes at most one spike to a volley, cells in higher visual areas may be shown to fire at most once upon stimulation, even in the presence of high population activity (e.g. LFP data recorded in parallel). This would indicate, but do not prove, the generation of volleys. Such *binary spiking* has already been detected in auditory cortex (DeWeese, Wehr, & Zador, 2003). Alternatively, reoccurring spike volleys may be unambiguously detected with single unit data (Schrader, Grün, Diesmann, & Gerstein, 2008), however requiring massively parallel recordings.

Our model predicts a difference between the initial phase of recognition, when an initial hypothesis must be established and the later phase of recognition, when a hypothesis was established

and is refined. Evidence for these two modes may be found in experiments with free-viewing animals. After saccades within an object, activity in the B-system, i.e. layers II/III, should be higher than after saccades to a new object.

4.10. Summary of results

The *Cortext* model successfully implements our hypothesized A- and B-systems with spiking neurons, thereby extending a study using analog neurons (Kupper et al., 2007). In the original concept (Koerner et al., 1999) the two infragranular layers V and VI are defined as a C-System that provides main output to subcortical structures, but also generates feedback to cortical areas (especially far reaching feedback over several hierarchy levels). As we use only short-range feedback to the next lower cortical area, we did not introduce the C-system in the current simulations. A non-spiking variant of the A-, B- and C-systems has been shown by a model by Kupper, Gewaltig, Knoblauch, Körner, and Körner (2008), we are currently incorporating the C-system into Cortext.

Our model is biologically grounded and demonstrates how an initial hypothesis is quickly generated via the feed-forward projections. This hypothesis is then fed back as prediction to the lower levels. During the feedback-mediated processing phase, the initial hypothesis can be refined or corrected. This works by switching-off those columns in lower levels that contribute to the hypothesis at the highest level. In this stage, columns which were previously suppressed by either lateral or forward inhibition have a chance to send their decisions to the next levels.

All processing phases in our *Cortext* model rely on spike times. During the feed-forward phase, a spike wave travels the hierarchy upwards. During the feedback-mediated phase, there is activity traveling in both directions. The feedback connections are either modulating (between areas) or inhibitory (between layers). This helps to sharpen the neuronal responses in time with the result that the spike-latency code is maintained also during the feedback-mediated phase.

The principal information carrier in our model is the relative response latency. While Delorme and Thorpe (2001) demonstrated how an object hypothesis can be generated with one spike per neuron, our *Cortext* model demonstrates how spike based processing can proceed beyond the first hypothesis.

The *Cortext* model, thus, provides a consistent interpretation for rapid recognition on the one hand and the abundance of feedback between areas on the other hand. The model also links the concept of cortical columns with the clearly visible cortical layers. Different processing modes are located in different cortical layers which are meaningfully arranged by vertical connections — the column. Each column thus implements a simple but powerful processing unit whose computing principle is repeated over the whole cortical surface.

Acknowledgments

We are grateful for the NEST Initiative in general and to Jochen Eppler in particular for help on the simulation environment NEST. We thank Rüdiger Kupper who assisted in the development of the network implementation and Heiko Wersing for comments on the model. We would also like to thank the anonymous reviewers for their constructive comments that helped to clarify and improve the original manuscript.

Appendix

Since Eq. (1) defines a linear model, the PSC amplitude J can be obtained by multiplying the amplitude of the desired PSP by a constant factor. Those are 165.44 pA/mV for excitatory PSCs and 26.6 pA/mV for inhibitory PSCs. These factors were obtained by numerically determining the maximum of a single PSP, whose time course is explicitly given by

$$\mathrm{PSP}(t) = \frac{eJ}{C_\mathrm{m}\tau_\mathrm{syn}} \times \left[\frac{-t\exp(-t/\tau_\mathrm{syn})}{1/\tau_\mathrm{syn}-1/\tau_\mathrm{m}} + \frac{\exp(-t/\tau_\mathrm{m})-\exp(-t/\tau_\mathrm{syn})}{(1/\tau_\mathrm{syn}-1/\tau_\mathrm{m})^2}\right]. \quad (8)$$

We use the correlation coefficient between the receptive field profiles to determine the overlap between all features (Table 5).

$$c_{L_1,L_2} = \frac{\sum_{i=1}^{n}(L_{1,i}-\mu_{L_1})(L_{2,i}-\mu_{L_2})}{\sqrt{\sum_{i=1}^{n}(L_{1,i}-\mu_{L_1})^2}\cdot\sqrt{\sum_{i=1}^{n}(L_{2,i}-\mu_{L_2})^2}} \quad (9)$$

where L_1 and L_2 are $\{0, 1\}$-vectors, representing the two letters, μ_{L_1} and μ_{L_2} their mean pixel values. n denotes the length of the vector. It is 64 for a single letter, and $m \cdot 64$ for a word of length m. Word similarity is defined by the average letter similarity and is used for lateral inhibition. As LGN relay cells are excitatory, correlation is clipped below zero.

Note that the correlation coefficient does not depend on the number of pixels in a letter. It will reach its maximum of 1 if the two letters are identical. As a result of the correlation coefficient between similar letters still being large (e.g. 0.82 between E and F), we potentiate the correlations by 3 in order to reach a steeper tuning characteristics of receptors, yielding p_i in Table 6. As we are also using diluted stimuli where pixels are randomly removed, the lack of pixels and therefore the decreased correlation must be compensated for. To this end, we normalize the correlation coefficients between the stimulus image and all receptor templates (23 coefficients total) such that the sum of all coefficients in a macrocolumn yields 1. We obtain the synaptic weight (PSP amplitude) by multiplying by a factor such that the sum of all weights is constant for all retinal positions (Table 6). Thus, the total charge that comes from any retina position in V1 is constant. The sum of all amplitudes of the resulting post-synaptic potentials that enter the sub-system of one retinal position is 8 mV. Words were generated by randomly selecting 80 three-letter syllables from English text. Therefore not all 26 letters were captured by this procedure. Syllables were then randomly arranged to form artificial words, each containing three syllables and nine letters. Words were chosen to have a correlation less than 0.85 to one another. The features in our model are listed as follows.

Letters: A, B, C, D, E, F, G, H, I, K, L, M, N, O, P, R, S, T, U, V, W, Y, Z

Syllables: AGE, AGO, AIR, ALL, AND, ANY, ART, BAM, BET, BOY, CAN, CRI, CUS, DEN, DIO, DON, ERN, ERS, FER, FOR, FRO, GAR, GAZ, GIV, HAD, HIM, HOW, LAU, LED, LIC, LID, LIT, LOW, MAK, MAS, MER, MET, MOD, MUR, NER, NEV, NIS, NOT, ODD, OUT, OWN, PAL, PAR, PEN, PEO, PLE, PLY, PRE, PUB, PUT, REL, RID, ROS, RUN, SAW, SAY, SIL, SIM, SON, SOR, SPO, STU, SUM, TEN, TER, THE, TLE, TOM, TOO, TRY, VAS, VEY, VUL, WHO, WHY

Words: AGEDENBET, AGEFROSON, AGESAWTLE, AGETOMLIC, AGOMAKNIS, AGOMODTRY, AIRERNPAL, AIRMODTEN, AIRPLYNIS, AIRPRESUM, AIRVASTOM, ALLBOYPUB, ALLGARPRE, ALLGIVGAR, ALLTERBAM, ANDHIMLOW, ANDMAKFOR, ANDNISMAS, ANDPUTPUT, ANDSONPUB, ANDSONSAW, ANYBETCRI, ANYBETPLY, ANYERNSUM, ANYFORMAK, ANYHADHAD, ANYHADTLE,

ANYMURPUT, ARTNERPAR, ARTPALNER, ARTPARTLE, ARTPLYTOM, ARTSAWPUT, ARTTENRUN, BAMCANSON, BAMLEDLOW, BAMRIDPEO, BETBOYMUR, BETGARROS, BETMETNIS, BETPRETLE, BETSILDIO, BETTRYGAR, BOYBETTOO, BOYCRIWHO, BOYHADPAR, CANFERDON, CANPLEHOW, CANRIDVEY, CRIBETCAN, CUSDENERN, CUSNERART, CUSPEOHOW, CUSVASSAY, DENHOWBAM, DENMASSUM, DENPUBSPO, DENPUBTHE, DENSILWHY, DENSONTEN, DONAGESPO, DONALLMUR, DONFERNOT, DONHOWDEN, DONMAKLED, DONMODTLE, DONNISSPO, DONNOTMET, ERNBAMFOR, ERNPALGAZ, ERNPENODD, ERNVEYPUB, ERNWHYHOW, ERSMAKVAS, FERANYHOW, FERHIMMUR, FERMAKSAW, FEROWNPAL, FERPUTLOW, FERSILCRI, FERSTUOUT, FORAGEDON, FORBOYPUT, FORRUNCAN, FRONISTER, FROPREMOD, GARERNODD, GAZDONODD, GIVDENSOR, GIVMODODD, HADFERSIM, HADNOTMOD, HADOWNPEN, HADSAWPAR, HIMDONGAR, HIMNEVPAR, HOWCANFER, LEDLOWPAL, LEDMODPEN, LEDNISFRO, LEDPLYCAN, LICDONVEY, LICPLYTOM, LICSAYHAD, LICSILGAZ, LIDBOYERN, LIDCRIFOR, LIDDONTOM, LIDNISTLE, LIDPENNIS, LIDPRETOM, LIDVULSUM, LITMODDEN, LITPALMAS, LITPLEBOY, LITSAWDIO, LITTERLAU, LITTOMPLE, LOWERNTHE, MAKERNMAK, MAKFORMUR, MAKPLEOWN, MAKTHEAND, MASMERHIM, MASSORSPO, MERAGOSON, MERARTFOR, MERLIDPEN, MERMASMAS, MEROUTMAS, MERPARVEY, MERPUBPAR, MERSONRID, METPUTBOY, METTLEBET, MODERNERN, MODLIDSPO, MODPLYERN, MODVULTOM, MURFORSAY, NERHADMAS, NERMASMOD, NERMERLAU, NERMODNER, NERMURRID, NERPLETOM, NERSONAGE, NEVDIOTLE, NEVERSDEN, NEVPALMET, NISGIVSOR, NISSTUTOO, NOTPUBPLE, NOTTHESIL, ODDFERFRO, ODDLOWBAM, ODDMAKSTU, ODDOUTGIV, ODDVEYHIM, OUTCANPUB, OUTNISLOW, OUTPLYERN, OWNBOYDON, OWNDENERN, OWNMAKGIV, OWNTENSPO, PALBETSAY, PALBOYTOM, PALHIMPLE, PALNERMAK, PALNEVGAR, PALTRYLOW, PARBAMPLY, PARMODPUT, PARSORPAR, PENCUSSOR, PENPLEOWN, PENPLYGAR, PENSORSUM, PENVASDON, PEOHADMER, PEOPLYBET, PLEHADWHY, PLEWHYGAR, PLYSAWPAL, PLYWHOERN, PREALLGIV, PREHADNER, PREHADSON, PRELICPAL, PREMAKTLE, PREPLEBOY, PREVASDEN, PREVEYDON, PUBBAMLIC, PUBCANSON, PUBMURNIS, PUBNERERS, PUBNEVPAR, PUBRIDDON, PUBRUNPLY, PUBSPOVAS, PUBTRYSUM, PUTFERLIT, PUTPREHOW, RELCUSPUT, RELHOWCRI, RELSAWMUR, RIDBETSAW, RIDLAUTHE, RIDLOWPAL, RIDNISMET, RIDPREHAD, RIDSUMNIS, RIDWHYVAS, ROSBOYPAL, ROSMERDEN, ROSNERCUS, ROSTOOGIV, RUNHADMET, RUNPALVAS, RUNSILBAM, SAWDONHOW, SAWDONPUT, SAWPENVUL, SAWPLYSUM, SAYERSBET, SAYHIMLIC, SAYPALREL, SAYPUBVUL, SAYVASNIS, SILPALSAW, SILPALSPO, SIMBETDEN, SIMCUSDIO, SIMMURWHY, SIMRUNMUR, SIMSAWPLY, SONPENBAM, SONROSNER, SONSORREL, SONSORSAW, SONTHEFER, SONTHEHOW, SONTOMPRE, SORMAKOUT, SORPLEPLE, SORSONAIR, SPOFORMOD, SPOHIMLID, STUAGEWHY, STUARTTLE, STUFORVEY, STUPLYBAM, SUMDENGAR, SUMNISSTU, SUMSORCUS, TENERNBET, TENFORCRI, TENFORPLY, TENMASMET, TENNOTMOD, TENPRESUM, TENROSSAW, TENSILNIS, TENSIMFRO, TENTHEVUL, TERDENMUR, TERGARNEV, TERGARROS, TERMETNER, THEMERPLE, THESORPUT, TLEBOYGAR, TLECANLAU, TLEHADPAR, TLEMAKFER, TLEMURCUS, TLEOUTTOM, TLEWHOSON, TOMLEDSIL, TOMLIDBAM, TOMSIMDON, TRYDONFOR, TRYGARSOR, TRYLOWPEN, TRYTHELID, TRYVASTEN, VASMAKFRO, VASMERPUB, VEYBETBAM, VEYDENPUB, VEYHOWPUT, VEYLAUMAS, VULHADPAL, VULLITLOW, VULNISAGE, WHOERNMER, WHOTOMSUM, WHYBETFOR

References

Adrian, E. (1928). *The basis of sensation: The action of the sense organs*. Hafner Publishing Co Ltd.

Barlow, H. B. (1972). Single units and sensation: A neuron doctrine for perceptual psychology? *Perception*, *1*(4), 371–394.

Barlow, H. B. (1994). What is the computational goal of the neocortex? In C. Koch, & J. L. Davis (Eds.), *Large-scale neuronal theories of the brain* (pp. 1–22). Cambridge, MA: MIT Press.

Blasdel, G. G., & Lund, J. S. (1983). Termination of afferent axons in macaque striate cortex. *Journal of Neuroscience*, *3*(7), 1389–1413.

Britten, K. H. (1998). Clustering of response selectivity in the medial superior temporal area of extrastriate cortex in the macaque monkey. *Vis Neuroscience*, *15*(3), 553–558.

Brodmann, K. (1909). *Localisation in the cerebral cortex*. London: Smith-Gordon, reprinted in 1994.

Bullier, J. (2001). Integrated model of visual processing. *Brain Research Reviews*, *36*, 96–107.

Butts, D., Weng, C., Jin, J., Yeh, C., Lesica, N., Alonso, J., et al. (2007). Temporal precision in the neural code and the timescales of natural vision. *Nature*, *449*(7158), 92–95.

Callaway, E. M., & Wiser, A. K. (1996). Contributions of individual layer 2–5 spiny neurons to local circuits in macaque primary visual cortex. *Vis Neuroscience*, *13*(5), 907–922.

Calvin, W. H., & Stevens, C. F. (1968). Synaptic noise and other sources of randomness in motoneuron interspike intervals. *Journal of Neurophysiology*, *31*(4), 574–587.

Casanova, M. (2005). *Neocortical modularity and the cell minicolumn*. Nova Biomedical Books.

Delorme, A., & Thorpe, S. J. (2001). Face identification using one spike per neuron: Resistance to image degradations. *Neural Networks*, *14*, 795–803.

DeWeese, M. R., Wehr, M., & Zador, A. M. (2003). Binary spiking in auditory cortex. *Journal of Neuroscience*, *23*(21), 7940–7949.

Eckhorn, R., Reitboeck, H. J., Arndt, M., & Dicke, P. (1990). Feature linking via synchronization among distributed assemblies: Simulations of results from cat visual cortex. *Neural Computing*, *2*, 293–307.

Eppler, J. M., Helias, M., Muller, E., Diesmann, M., & Gewaltig, M. O. (2008). PyNEST: A convenient interface to the NEST simulator. *Front Neuroinformatics*, *2*, 12–12.

Felleman, D., & Van Essen, D. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, *1*(1), 1–47.

Ferster, D., & Lindström, S. (1983). An intracellular analysis of geniculo-cortical connectivity in area 17 of the cat. *Journal of Physiology*, *342*, 181–215.

Fitzpatrick, D., Lund, J. S., & Blasdel, G. G. (1985). Intrinsic connections of macaque striate cortex: afferent and efferent connections of lamina 4c. *Journal of Neuroscience*, *5*(12), 3329–3349.

Furl, N., van Rijsbergen, N., Treves, A., Friston, K., & Dolan, R. (2007). Experience-dependent coding of facial expression in superior temporal sulcus. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(33), 13485.

Gewaltig, M.-O., & Diesmann, M. (2007). NEST (Neural Simulation Tool). *Scholarpedia*, *2*(4), 1430.

Gewaltig, M.-O., Diesmann, M., & Aertsen, A. (2001). Propagation of cortical synfire activity: Survival probability in single trials and stability in the mean. *Neural Networks*, *14*, 657–673.

Grossberg, S., & Raizada, R. D. (2000). Contrast-sensitive perceptual grouping and object-based attention in the laminar circuits of primary visual cortex. *Vision Research*, *40*(10–12), 1413–1432.

Grossberg, S., & Versace, M. (2008). Spikes, synchrony, and attentive learning by laminar thalamocortical circuits. *Brain Research*, *1218*, 278–312.

Gur, M., & Snodderly, D. M. (2008). Physiological differences between neurons in layer 2 and layer 3 of primary visual cortex (V1) of alert macaque monkeys. *Journal of Physiology*, *586*(9), 2293–2306.

Harrison, L. M., Stephan, K. E., Rees, G., & Friston, K. J. (2007). Extra-classical receptive field effects measured in striate cortex with fMRI. *Neuroimage*, *34*(3), 1199–1208.

Heil, P. (1997). Auditory cortical onset responses revisited. I. first-spike timing. *Journal of Neurophysiology*, *77*(5), 2616–2641.

Hirsch, J. A., & Martinez, L. M. (2006). Laminar processing in the visual cortical column. *Current Opinion in Neurobiology*, *16*(4), 377–384.

Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron*, *36*, 791–804.

Horton, J. C., & Adams, D. L. (2005). The cortical column: A structure without a function. *Philosophical Transactions of Royal Society of London Series B Biological Science*, *360*(1456), 837–862.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Neurophysiology*, *195*(1), 215–243.

Hübener, M., Shoham, D., Grinvald, A., & Bonhoeffer, T. (1997). Spatial relationships among three columnar systems in cat area 17. *Journal of Neuroscience*, *17*(23), 9270–9284.

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *National Reviews in Neuroscience*, *2*(3), 194–203.

Jack, J. J. B., Redman, S. J., & Wong, K. (1981). The components of synaptic potentials evoked in cat spinal motoneurones by impulses in single group Ia afferents. *Journal of Physiology (Lond)*, *321*, 65–96.

Johansson, R., & Birznieks, I. (2004). First spikes in ensembles of human tactile afferents code complex spatial fingertip events. *Natural Neuroscience*, *2*(7), 170–177.

Kiebel, S. J., Daunizeau, J., & Friston, K. J. (2008). A hierarchy of time-scales and the brain. *PLoS Computational Biology*, *4*(11).

Koerner, E., Gewaltig, M.-O., Koerner, U., Richter, A., & Rodemann, T. (1999). A model of computation in neocortical architecture. *Neural Networks*, *12*, 989–1005.
Koerner, E., Tsujino, H., & Masutani, T. (1997). A cortical-type modular neural network for hypothetical reasoning. *Neural Networks*, *10*(5), 791–814.
Kupper, R., Gewaltig, M.-O., Knoblauch, A., Körner, U., & Körner, E. (2008). From neurons to cortex: A multi-level approach to understanding the brain. In H. A. Svensson (Ed.)., *Neurocomputing research developments* (pp. 39–80). Nova Science Publishers, Inc., Chapter 2.
Kupper, R., Knoblauch, A., Gewaltig, M.-O., Koerner, U., & Koerner, E. (2007). Simulations of signal flow in a functional model of the cortical column. *Neurocomputing*, *70*(10–12), 1711–1716.
Lomber, S. G., & Payne, B. R. (2000). Translaminar differentiation of visually guided behaviors revealed by restricted cerebral cooling deactivation. *Cerebral Cortex*, *10*(11), 1066–1077.
Lorente de No, R. (1949). Cerebral cortex: Architecture, intracortical connections, motor projections. In J. F. Fulton (Ed.)., *Physiology of the nervous system* (pp. 288–330). Oxford: Oxford University Press, Chapter 15.
MacLeod, K., & Laurent, G. (1996). Distinct mechanisms for synchronization and temporal patterning of odor-encoding neural assemblies. *Science*, *274*(5289), 976–979.
Markram, H., Toledo-Rodriguez, M., Wang, Y., Gupta, A., Silberberg, G., & Wu, C. (2004). *Interneurons of the Neocortical Inhibitory System*, *5*(10), 793–807.
Mountcastle, V. (2003). Introduction. *Cerebral Cortex*, *13*, 2–4.
Mountcastle, V. B. (1957). Modality and topographic properties of single neurons of cat's somatic sensory cortex. *Journal of Neurophysiology*, *20*(4), 408–434.
Mountcastle, V. B. (1997). The columnar organization of the neocortex. *Brain*, *120*(4), 701–722.
Mumford, D. (1994). Neuronal architectures for pattern-theoretic problems. In C. Koch, & D. J. L (Eds.), *Large-scale neuronal theories of the brain* (pp. 125–152). Cambridge, MA: MIT Press.
Murray, S., Kersten, D., Olshausen, B., Schrater, P., & Woods, D. (2002). Shape perception reduces activity in human primary visual cortex. *Proceedings of the National Academy of Sciences*, *99*(23), 15164–15169.
Nassi, J. J., & Callaway, E. M. (2007). Specialized circuits from primary visual cortex to V2 and area MT. *Neuron*, *55*(5), 799–808.
Nordlie, E., Gewaltig, M.-O., & Plesser, H. E. (2009). Towards reproducible descriptions of neuronal network models. *PLoS Computational Biology*, *5*(8), e1000456.
Olshausen, B. A., Anderson, C. H., & van Essen, D. C. (1993). A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. *Journal of Neuroscience*, *13*(11), 4700–4719.
Osborne, L. C., Bialek, W., & Lisberger, S. G. (2004). Time course of information about motion direction in visual area MT of macaque monkeys. *Journal of Neuroscience*, *24*(13), 3210–3222.
Papoulis, A. (1991). *Probability, random variables, and stochastic processes* (3rd ed.). Boston, Massachusetts: McGraw-Hill.
Peters, A., & Jones, E. G. (Eds.). (1984). *Cerebral cortex: Vol. 1. Cellular components of the cerebral cortex*. New York: Plenum.
Porter, J. T., Johnson, C. K., & Agmon, A. (2001). Diverse types of interneurons generate thalamus-evoked feedforward inhibition in the mouse barrel cortex. *Journal of Neuroscience*, *21*(8), 2699–2710.
Raizada, R., & Grossberg, S. (2003). Towards a theory of the laminar architecture of cerebral cortex: Computational clues from the visual system. *Cerebral Cortex*, *13*, 100–113.
Rakic, P. (2008). Confusing cortical columns. *Proceedings of the National Academy of Sciences*, *105*(34), 12099.
Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87.
Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, *2*(11), 1019–1025.
Rockland, K. S., & Ichinohe, N. (2004). Some thoughts on cortical minicolumns. *Experimental Brain Research*, *158*(3), 265–277.
Rockland, K. S., & Van Hoesen, G. W. (1994). Direct temporal-occipital feedback connections to striate cortex (V1) in the macaque monkey. *Cerebral Cortex*, *4*(3), 300–313.
Rockland, K. S., & Virga, A. (1989). Terminal arbors of individual "feedback" axons projecting from area V2 to V1 in the macaque monkey: A study using immunohistochemistry of anterogradely transported phaseolus vulgaris-leucoagglutinin. *Journal of Comparative Neurology*, *285*(1), 54–72.
Saul, A. (2008). Lagged cells in alert monkey lateral geniculate nucleus. *Visual Neuroscience*, *25*(5–6), 647–659.
Sawatari, A., & Callaway, E. M. (2000). Diversity and cell type specificity of local excitatory connections to neurons in layer 3b of monkey primary visual cortex. *Neuron*, *25*(2), 459–471.
Schrader, S., Grün, S., Diesmann, M., & Gerstein, G. (2008). Detecting synfire chain activity using massively parallel spike train recording. *Journal of Neurophysiology*, *100*, 2165–2176.
Sherman, S. M. (2007). The thalamus is more than just a relay. *Current Opinion in Neurobiology*, *17*(4), 417–422.
Sherrington, C. (1941). Man on his nature. In *The Gifford Lectures, Edinburgh 1937–38*. Cambridge: Cambridge University Press.
Shmuel, A., Korman, M., Sterkin, A., Harel, M., Ullman, S., Malach, R., et al. (2005). Retinotopic axis specificity and selective clustering of feedback projections from V2 to V1 in the owl monkey. *Journal of Neuroscience*, *25*(8), 2117–2131.
Steriade, M., Jones, E. G., & McCormick, D. A. (1997). *Thalamus*. New York: Elsevier.
Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*(6), 520–522.
Thorpe, S. J. (1990). Spike arrival times: A highly efficient coding scheme for neural networks. In R. Eckmiller, G. Hartmann, & G. Hauske (Eds.), *Parallel processing in neural systems* (pp. 91–94). Amsterdam: North-Holland.
Thorpe, S. J., Delorme, A., & Van Rullen, R. (2001). Spike-based strategies for rapid processing. *Neural Networks*, *14*(6–7), 715–725.
Tuckwell, H. C. (1988). The Lapique model of the nerve cell. In *Introduction to theoretical neurobiology Vol. 1* (pp. 85–123). Cambridge: Cambridge University Press, Chapter 3.
Van Essen, D. C., Newsome, W. T., Maunsell, J. H., & Bixby, J. L. (1986). The projections from striate cortex (V1) to areas V2 and V3 in the macaque monkey: Asymmetries, areal boundaries, and patchy connections. *Journal of Comparative Neurology*, *244*(4), 451–480.
Van Rullen, R., & Thorpe, S. J. (2002). Surfing a spike wave down the ventral stream. *Visual Research*, *42*, 2593–2615.
Von Economo, K., & Koskinas, G. N. (1925). *Die Cytoarchitektonik der Hirnrinde des erwachsenen Menschen*. Berlin: Springer-Verlag.
Wang, Q., & Burkhalter, A. (2007). Area map of mouse visual cortex. *Journal of Comparative Neurology*, *502*(3), 339–357.
Wang, W., Jones, H. E., Andolina, I. M., Salt, T. E., & Sillito, A. M. (2006). Functional alignment of feedback effects from visual cortex to thalamus. *Nature Neuroscience*, *9*(10), 1330–1336.
Watt, R. J., & Phillips, W. A. (2000). The function of dynamic grouping in vision. *Trends in Cognitive Science*, *4*(12), 447–454.
Wersing, H., & Körner, E. (2003). Learning optimized features for hierarchical models of invariant object recognition. *Neural Computing*, *15*(7), 1559–1588.
White, E. L. (1989). *Cortical circuits: Synaptic organization of the cerebral cortex—Structure, function, and theory*. Boston: Birkhauser.
Zhu, S.-C., & Mumford, D. (2006). A stochastic grammar of images. *Foundations and Trends in Computer Graphics and Vision*, *2*(4), 259–362.