Self-Organization of Spiking Neural Networks for Visual Object Recognition

Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften
(Dr. rer. nat.)

dem Fachbereich Biologie
der Philipps-Universität Marburg
vorgelegt von

Frank Michler
aus Karl-Marx-Stadt

Marburg/Lahn 2019

Vom Fachbereich Biologie der Philipps-Universität Marburg
(Hochschulkennziffer 1180) als Dissertation am 02.12.2019 angenommen.

Erstgutachter: Prof. Dr. Thomas Wachtler (Ludwig-Maximilians-Universität München)
Zweitgutachter: Prof. Dr. Uwe Homberg (Philipps-Universität Marburg)

Tag der mündlichen Prüfung: 19.12.2019

Eidesstattliche Erklärung

Ich, Frank Michler, versichere, dass ich meine Dissertation mit dem Titel

„Self-Organization of Spiking Neural Networks for Visual Object Recognition“

selbständig, ohne unerlaubte Hilfe angefertigt und mich dabei keiner anderen als der von mir ausdrücklich bezeichneten Quellen und Hilfen bedient habe.

Die Dissertation wurde in der jetzigen oder einer ähnlichen Form noch bei keiner anderen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.

Unterschrift:

Datum:

Zusammenfassung

Unser visuelles System hat zum einen die Fähigkeit, sehr ähnliche Objekte zu unterscheiden. Zum anderen können wir dasselbe Objekt wiedererkennen, obwohl sich seine Abbildung auf der Netzhaut aufgrund des Blickwinkels, des Abstandes oder der Beleuchtung stark unterscheiden kann. Diese Fähigkeit, dasselbe Objekt in unterschiedlichen Netzhaut-Bildern wiederzuerkennen, wird als *invariante Objekterkennung* bezeichnet und ist noch nicht sofort nach der Geburt verfügbar. Sie wird erst durch Erfahrung mit unserer visuellen Umwelt erlernt.

Häufig sehen wir verschiedene Ansichten desselben Objektes in einer zeitlichen Abfolge, zum Beispiel wenn es sich selbst bewegt oder wir es in unserer Hand bewegen, während wir es betrachten. Dies erzeugt zeitliche Korrelationen zwischen aufeinander folgenden Netzhaut-Bildern, die dazu verwendet werden können, verschiedene Ansichten desselben Objektes miteinander zu assoziieren. Theoretiker vermuten daher, dass eine synaptische Lernregel mit einer eingebauten Gedächtnisspur (englisch: *trace rule*) dazu verwendet werden kann, invariante Objektrepräsentationen zu lernen.

In dieser Dissertation stelle ich Modelle für impulskodierende neuronale Netze (englisch: *spiking neural networks*) zum Lernen invarianter Objektrepräsentationen vor, die auf folgenden Hypothesen beruhen:

1. Anstelle einer synaptischen *trace rule* kann persistente Spike-Aktivität von vernetzten Neuronengruppen als eine Gedächtnis-Spur für Invarianz-Lernen dienen.

2. Kurzreichweitige laterale Verbindungen ermöglichen das Lernen von selbst organisierenden topographischen Karten, welche neben räumlichen auch zeitliche Korrelationen abbilden.

3. Wird ein solches Netzwerk mit Bildern von kontinuierlich rotierenden Objekten trainiert, so kann es Repräsentationen lernen, in denen Ansichten desselben Objekts benachbart sind. Derartige Objekttopographien können invariante Objekterkennung ermöglichen.

4. Das Lernen von Repräsentationen sehr ähnlicher Muster kann durch anpassungsfähige inhibierende Feedback-Verbindungen ermöglicht werden.

Die in Kapitel 3.1 vorgestellte Studie legt die Implementierung eines impulskodierenden neuronalen Netzes dar, an welchem die ersten drei Hypothesen überprüft wurden. Das Netzwerk wurde mit Stimulus-Sets getestet, in denen die Stimuli in zwei Merkmalsdimensionen so angeordnet waren, dass sich der Einfluss von zeitlichen und räumlichen Korrelationen auf die gelernten topographischen Karten trennen ließ. Die entstandenen topographischen Karten wiesen Muster auf, welche von der zeitlichen Reihenfolge der beim Lernen präsentierten Objektansichten abhingen. Unsere Ergebnisse zeigen, dass durch die Zusammenfassung der neuronalen Aktivitäten aus einer lokalen Nachbarschaft der topographischen Karten invariante Objekterkennung ermöglicht wird.

Das Kapitel 3.2 beschäftigt sich mit der vierten Hypothese. In dieser Publikation wurden die Untersuchungen dazu beschrieben, wie adaptive Feedback-Inhibition (AFI) die Fähigkeit eines Netzwerkes verbessern kann, zwischen sehr ähnlichen Mustern zu unterscheiden. Die Ergebnisse zeigen, dass mit AFI schneller stabile Muster-Repräsentationen gelernt wurden und dass Muster mit einem höheren Grad an Ähnlichkeit unterschieden werden konnten als ohne AFI.

Die Ergebnisse von Kapitel 3.1 zeigen eine funktionale Rolle für topographische Objekt-Repräsentationen auf, welche aus dem inferotemporalen Kortex bekannt sind, und erklären, wie diese sich herausbilden können. Das AFI-Modell setzt einen Aspekt der *Predictive Coding*-Theorie um: die Subtraktion einer Vorhersage vom tatsächlichen Input eines Systems. Die erfolgreiche Implementierung dieses Konzepts in einem biologisch plausiblen Netzwerk impulskodierender Neuronen zeigt, dass das *Predictive Coding*-Prinzip in kortikalen Schaltkreisen eine Rolle spielen kann.

Abstract

On one hand, the visual system has the ability to differentiate between very similar objects. On the other hand, we can also recognize the same object in images that vary drastically, due to different viewing angle, distance, or illumination. The ability to recognize the same object under different viewing conditions is called *invariant object recognition*. Such object recognition capabilities are not immediately available after birth, but are acquired through learning by experience in the visual world.

In many viewing situations different views of the same object are seen in a temporal sequence, e.g. when we are moving an object in our hands while watching it. This creates temporal correlations between successive retinal projections that can be used to associate different views of the same object. Theorists have therefore proposed a synaptic plasticity rule with a built-in memory trace (*trace rule*).

In this dissertation I present spiking neural network models that offer possible explanations for learning of invariant object representations. These models are based on the following hypotheses:

1. Instead of a synaptic trace rule, persistent firing of recurrently connected groups of neurons can serve as a memory trace for invariance learning.
2. Short-range excitatory lateral connections enable learning of self-organizing topographic maps that represent temporal as well as spatial correlations.
3. When trained with sequences of object views, such a network can learn representations that enable invariant object recognition by clustering different views of the same object within a local neighborhood.
4. Learning of representations for very similar stimuli can be enabled by adaptive inhibitory feedback connections.

The study presented in chapter 3.1 details an implementation of a spiking neural network to test the first three hypotheses. This network was tested with stimulus sets that were designed in two feature dimensions to separate the impact of temporal and spatial correlations on learned topographic maps. The emerging topographic maps showed patterns that were dependent on the temporal order of object views during training. Our results show that pooling over local neighborhoods of the topographic map enables invariant recognition.

Chapter 3.2 focuses on the fourth hypothesis. There we examine how the *adaptive feedback inhibition* (AFI) can improve the ability of a network to discriminate between very similar patterns. The results show that with AFI learning is faster, and the network learns selective representations for stimuli with higher levels of overlap than without AFI.

Results of chapter 3.1 suggest a functional role for topographic object representations that are known to exist in the inferotemporal cortex, and suggests a mechanism for the development of such representations. The AFI model implements one aspect of *predictive coding*: subtraction of a prediction from the actual input of a system. The successful implementation in a biologically plausible network of spiking neurons shows that predictive coding can play a role in cortical circuits.

List of Abbreviations

AFI	**A**daptive **F**eedback **I**nhibition
AMPA	α-**a**mino-3-hydroxy-5-**m**ethyl-4-isoxazole**p**ropionic **a**cid
AP	**A**ction **P**otential
CNN	**C**onvolutional **N**eural **N**etwork
CT	**C**ontinuous **T**ransformation
EPSC	**E**xcitatory **P**ost-**S**ynaptic **C**urrent
EPSP	**E**xcitatory **P**ost-**S**ynaptic **P**otential
GABA	**G**amma-**A**mino**b**utyric **A**cid
IPSC	**I**nhibitory **P**ost-**S**ynaptic **C**urrent
IPSP	**I**nhibitory **P**ost-**S**ynaptic **P**otential
LIF	**L**eaky **I**ntegrate-and-**F**ire
LTD	**L**ong **T**erm **D**epression
LTP	**L**ong **T**erm **P**otentiation
NMDA	**N**-**m**ethyl-**D**-**a**spartate
NMDAR	**N**-**m**ethyl-**D**-**a**spartate **R**eceptor
SNN	**S**piking **N**eural **N**etwork
SOM	**S**elf **O**rganizing **M**ap
STDP	**S**pike **T**iming **D**ependent **P**lasticity
WTA	**W**inner-**T**ake-**A**ll

Contents

Eidesstattliche Erklärung **iii**

Zusammenfassung **v**

Abstract **vii**

1 Introduction **1**
- 1.1 Vision in Biological and Artificial Systems 1
- 1.2 Learning 2
- 1.3 Neural Network Models for Object Recognition 3
- 1.4 Hypotheses and Objectives 8

2 Methodological Background: Simulating Neural Networks **11**
- 2.1 Modeling: The Art of Simplification 11
- 2.2 Model Neurons 11
- 2.3 Layers 15
- 2.4 Synaptic Transmission 15
- 2.5 Cellular Mechanisms of Neural Plasticity 17
- 2.6 Synaptic Learning Rules 18
- 2.7 Competition: The Winner Takes it All 19

3 Publications **21**
- 3.1 Spatiotemporal Correlations and Topographic Maps 21
- 3.2 Adaptive Feedback Inhibition 37

4 Discussion **51**
- 4.1 Invariant Object Recognition 52
- 4.2 Trace Learning in Spiking Neural Networks 53
- 4.3 Sustained Intrinsic Activity 54
- 4.4 Empirical Evidence for the Role of Temporal Contiguity 55
- 4.5 Adaptive Feedback Inhibition and Predictive Coding 56
- 4.6 Combining AFI and Topographic Map Learning 58
- 4.7 Why Study Spiking Neural Networks? 59
- 4.8 Conclusion 60

Bibliography **61**

Chapter 1

Introduction

1.1 Vision in Biological and Artificial Systems

Vision is highly important in our daily life, which is also reflected in our language (San Roque et al., 2015). Vision is not just about detecting light, but about reconstructing and interpreting our environment from the light patterns that activate photoreceptors in the retina. Therefore, understanding the principles of visual processing in the brain significantly contributes to our understanding of the human brain itself.

In recent years, test projects with self driving cars on public roads have been started (Waldrop, 2015; Zoellick et al., 2019). This was made possible by the progress of modern computer vision systems, which use multi layered architectures with a processing hierarchy that is inspired by insights gained from studying the human and mammalian visual system (Chen et al., 2019). This exemplifies how empirical and theoretical neuroscience research has translated into technical solutions that can improve our lives. Yet, there are still many unsolved problems, such as learning of object representations from a continuous stream of inputs, without relying on training with huge labeled datasets. New insights into the way our brain achieves visual object recognition can trigger further progress.

Many of the computer vision systems used in cameras, self driving cars, or at large internet companies, are trained in a supervised way using huge databases of images that have been categorized and labeled manually by humans. In contrast, humans do not need a teacher to learn basic object recognition. We learn to recognize faces and objects through experience with the visual world (Ruff, Kohler, and Haupt, 1976). Temporal contiguity can provide cues that can be used in neural networks to associate different views of the same object. Some studies have already established that this principle plays a role in humans (Wallis and Bülthoff, 2001) and animals (Wood and Wood, 2018). But how exactly the brain makes use of temporal cues is still unknown.

The basic computational units in technical solutions for object recognition represent neural activity as an average firing rate, thereby abstracting away individual action potentials (APs, also called *spikes*). This approach simplifies computations and has lead to huge progress, because it enables simulations with large numbers of neurons. But information processing in the brain probably also relies on mechanisms that make use of the precise timing of individual spikes (Gollisch and Meister, 2008).

In this dissertation I will present two studies that address complementary problems of visual object recognition. The first study addresses the question of how objects can be recognized despite large variations of their retinal projections due to conditions like viewing angle, distance, and illumination (Michler, Eckhorn, and Wachtler, 2009, see section 3.1). The second study addresses how objects can be differentiated from each other despite large similarities (Michler, Wachtler, and Eckhorn, 2006, see section 3.2). In both studies we developed spiking neural networks that adjust their internal connections through unsupervised learning.

In the following sections of this introduction I will provide some background on the relationship of vision and learning, and neural network models for object recognition in order to explain the objectives and hypotheses of this dissertation.

1.2 Learning

Visual Perception Depends on Learning

When we look around, we easily recognize the face of a friend we want to talk to, or an apple we want to eat. This happens within a fraction of a second (Thorpe, Fize, and Marlot, 1996). But we are not born with these abilities. While non-mammals have innate abilities to navigate (Homberg et al., 2011), detect food (Lettvin et al., 1959), or recognize potential mates and enemies (Land, 1969; Dorosheva, Yakovlev, and Reznikova, 2011), many aspects of mammal and human vision are learned.

Even the fundamental ability to discriminate between horizontal and vertical edges relies on experience with the visual world, as was demonstrated by the groundbreaking experiments of Hubel and Wiesel (1970) and Blakemore and Cooper (1970) with cats.

For kittens it was shown that depriving visual input to one eye during a critical period in their development (first three months after birth) drastically reduced the response of neurons in the striate cortex to input from that eye (Hubel and Wiesel, 1970).

Neurons in the striate cortex of cats selectively respond to visual edges with a specific orientation (Hubel and Wiesel, 1962). In normal cats, optimal orientation is uniformly distributed. However, when kittens were exclusively exposed to vertical edges during the first five months of their lives, fewer cells were found with an optimal orientation perpendicular to the orientations the kittens had been exposed to. Also, their ability to see horizontal contours was drastically impaired (Blakemore and Cooper, 1970).

A reductionist approach leads to the question of how selectivity for the orientation of edges or representations of visual objects can emerge through learning on a cellular level.

Synaptic Plasticity and Hebbian Learning

How can experience induce long lasting changes of our perception and behavior? Cajal (1894) was the first to suggest that changes in the synapse are the cellular basis for learning.

Studies on hippocampus fibers have revealed experimental proof for Cajal's prediction. After repetitive stimulation, Bliss and Lømo (1973) found long lasting potentiation of excitatory postsynaptic potential (EPSP) amplitudes. This is referred to as long term potentiation (LTP). With prolonged low frequency stimulation hippocampal synapses also show a form of long-lasting synaptic depression (long term depression, LTD). Hebb (1949) postulated a principle explaining how these changes take place:

> "When an axon of cell A is near enough to excite cell B or repeatedly or consistently takes part in firing it, some growth or metabolic change

> takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased." (Donald Hebb, 1949)

Evidence for such learning mechanisms has been found by Markram et al. (1997), using whole-cell voltage recordings from neighboring neurons. They showed that coincidence of postsynaptic action potentials (APs) and unitary EPSPs induce changes in EPSPs. Bi and Poo (1998) measured how LTP and LTD occur depending on the precise timing of pre- and postsynaptic APs. This spike timing dependent plasticity (STDP) fulfills Hebb's postulate and enables synapses to work as causality detectors. The cellular mechanisms underlying STDP will be reviewed in more detail in section 2.5.

After examining the cellular level, I will now turn to the question of how networks of neurons and synapses exhibiting Hebbian plasticity can learn to represent and recognize visual objects. Since it is difficult to imagine how thousands of cells interact, computer simulations of neural networks can help to gain insights into the emergence of higher level properties, like view point invariance, from lower level processes.

1.3 Neural Network Models for Object Recognition

Standard Model for Pattern Recognition

When we see an object, it reflects photons that hit the retina, where photoreceptors and ganglion cells transform the information into patterns of neural activity. Thus, for the brain object recognition is a problem of pattern recognition. Many modern neural networks build upon the concepts first developed in the *perceptron* model (Rosenblatt, 1958). In its basic form it consists of three groups of neurons: a "projection area" A_I, which receives retinal input, an "association area" A_{II}, and "response cells" $R_1, R_2, ..., R_n$, which represent the output of the model. Such groups of neurons that share a functional role and are in the same level of a processing hierarchy are also often referred to as *layers* (Figure 1.1).

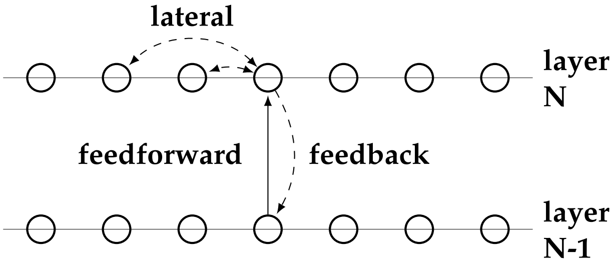


Figure 1.1: Feedforward, feedback, and lateral connections. Adapted from Intrator and Edelman (1997). Hierarchical neural networks are structured in layers. Connections from lower to higher levels are called *feedforward*, while *feedback* connections project from a higher level layer back to a lower level layer. *Lateral* connections connect neurons within a layer.

The activity value of a neuron is calculated from a weighted sum of the activity of its inputs. The strength of a connection is therefore often referred to as a *weight*, corresponding to the synaptic efficacy of biological neurons. In the perceptron-model of Rosenblatt (1958), weights of feedforward connections from A_I to A_{II} and from A_{II} to response cells are adjusted according to an error signal: the difference between desired and actual output. For perceptrons with more layers (*multi layer perceptrons*, MLPs), Werbos introduced a learning algorithm in which the error signal is propagated backwards through the processing hierarchy to update weights (Werbos, 1975; Werbos, 1990). This algorithm is called *backpropagation* and is a form of *supervised learning*, because the desired output of the network must be known beforehand to control the learning process. MLPs have been successfully applied to solve complex pattern recognition problems (e.g. recognition of handwritten characters, Jameel and Kumar, 2018).

To adjust weights in an unsupervised manner, a Hebbian plasticity rule can be used to calculates weight changes from the activity of pre- and postsynaptic neurons (section 1.2). The rule allows neurons to adjust the weights of their afferent synapses to match the activity pattern of presynaptic input neurons whenever a postsynaptic spike occurs. *Lateral* inhibition (Figure 1.1) can enhance activity differences and thereby prevent that all neurons learn the same pattern (Grossberg, 1973). When only one neuron within a layer is allowed to fire this is called a *winner-take-all* (WTA) network.

However, object recognition is more than just pattern recognition, since multiple input patterns can represent the same object. The challenge to generalize across multiple patterns and classify them as the same object is a fundamental problem in biological and machine vision (Simard et al., 1991; Zhang, 2019). Gibson (1966) hypothesized that ”constant perception depends on the ability of the individual to detect the invariants.”

Complex Cells as a Model for Invariance

When we watch a moving object, or make an eye movement between different points on an object, the retinal activity pattern changes drastically. To recognize the object, an internal representation is needed that is invariant with respect to these changes. Hubel and Wiesel (1962) have observed response properties in the cat visual cortex that could provide a basis for position invariance. Whereas some cells selectively responded to visual edges of a certain orientation at a specific position in the visual field ("simple cells"), other cells showed a similar selectivity for orientation, but responded equally strong for edges at different positions ("complex cells").

A model to explain these response properties was proposed by Hubel and Wiesel (1962): complex cells receive input from simple cells that are selective for the same orientation (S1 to C1 connections in Figure 1.2). Fukushima (1980) has proposed that this principle of simple and complex cells is repeatedly applied within the hierarchy of the visual system. Fukushima’s *neocognitron* model consists of a hierarchy of modules, each of which is comprised of a simple cell and a complex cell layer.

Riesenhuber and Poggio (1999) adopted this concept in their HMAX model: complex cells are "pooling" from groups of simple cells by performing a "MAX" operation on the output of simple cells with the same orientation preference (the output of the complex cell is equal to the maximum output of a set of simple cells with the same orientation but different position). The next layer in the hierarchy consists of "composite feature cells" (S2 cells in Figure 1.2), which perform a weighted sum over the output of complex cells. Their output is then pooled again to achieve tolerance for some transformations of the composite features.

The same principle is used in *Convolutional Neural Networks* (CNNs or ConvNets) which use alternating convolution and pooling layers (LeCun et al., 1998; LeCun, Bengio, and Hinton, 2015). Whereas many models of the visual system share the

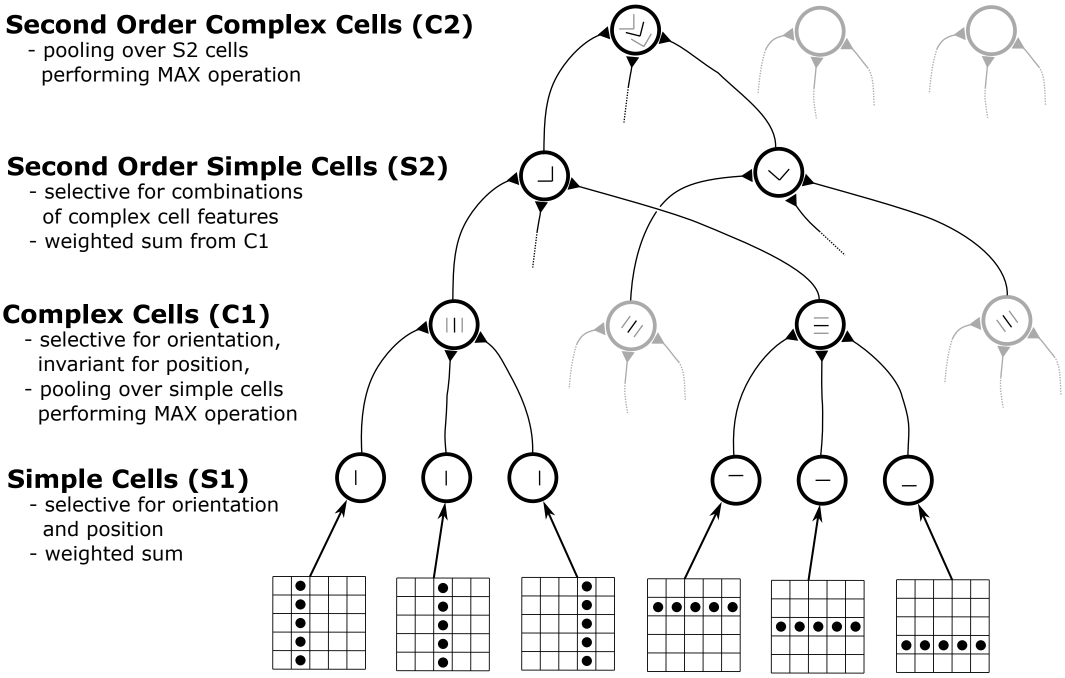


Figure 1.2: Sketch of the HMAX model (Riesenhuber and Poggio, 1999). Simple cells (S1) are selective for the precise position oriented edges, calculating a weighted sum across their inputs, cells in the lateral geniculate nucleus (LGN) with linearly aligned receptive field centers. Complex cells (C1) pool over simple cells with the same orientation preference but different positions (as proposed by Hubel and Wiesel, 1962). Pooling can be achieved with a MAX operation: output of a C1 cell is equal to the maximum output of its input S1 cells. Second order simple cells (S2) receive input from C1, performing a weighted sum operation. Therefore, they are selective for specific combinations of C1 features. Second order complex cells (C2) pool over S2 cells, thereby achieving higher order invariance. The example shows a C2 cell selective for corners of a specific opening angle and invariant with respect to the rotation angle.

concept of simple and complex cells, they differ in the way the underlying connectivity is established, and it still remains unknown how representations for invariant object recognition are learned in the brain.

Supervised vs Unsupervised Learning

How can a network determine which pattern detectors belong together as representations of the same object? Supervised learning using the backpropagation (BP) algorithm (Werbos, 1990; Rumelhart, Hinton, and Williams, 1986) has been applied successfully to solve complex object recognition problems, even surpassing human performance in specific classification tasks (He et al., 2015). For these algorithms, huge sets of training stimuli are needed for which the correct classification is already known (images are "labeled"). Each item of a training data set is presented to the network, and the difference between the correct output and the actual output is used as an error signal to adjust weights of internal synapses. The error signal is propagated backwards through the hierarchy of layers from the output layer to the input layer, hence the name "backpropagation".

While this approach is viable for technical systems, humans and animals do not learn object recognition by relying on pre-classified stimulus sets. Further, a number of issues have been raised that make backpropagation biologically unplausible (Bengio et al., 2015).

The brain likely uses unsupervised learning mechanisms to build internal representations for object recognition that rely only on the interactions within the genetically predetermined network architecture, mechanisms for synaptic plasticity, and experience with real world input.

Fukushima proposed a mechanism for unsupervised learning of simple cell connections (Fukushima, 1975; Fukushima, 1980). In this model, one unit with the strongest activation within a group of competing units (single cells receiving input from the same position of the visual field) is selected for learning after each presentation of an input pattern. Weights are adjusted in proportion to the activity of afferent units. This is a *winner-take-all* (WTA) algorithm and can be implemented biologically with a combination of lateral inhibition and Hebbian plasticity. This learning mechanism is based on similarity. Simple cells with afferent connections that most closely resemble the current input pattern win the competition, and weights of incoming connections belonging to the current input pattern are increased. However, for learning invariant representations this is not optimal, as I will explain in the next section.

Invariant Representations based on Temporal Proximity

To recognize objects under different viewing conditions, relying only on spatial correlations (i.e. similarity) is not sufficient: The frontal and profile views of one face result in very different retinal projections. On the other hand, frontal views of different faces can be very similar. Any neural learning mechanism that solely relies on similarity would therefore group images of different faces from the same viewing angle together, instead of associating different views of the same face.

In many natural viewing situations such as moving around while watching an object, or examining an object in our hands while rotating it, we see different views of that object successively (Figure 1.3). Therefore, temporal proximity can provide a cue for grouping retinal input patterns that belong to the same object. Földiák (1991)

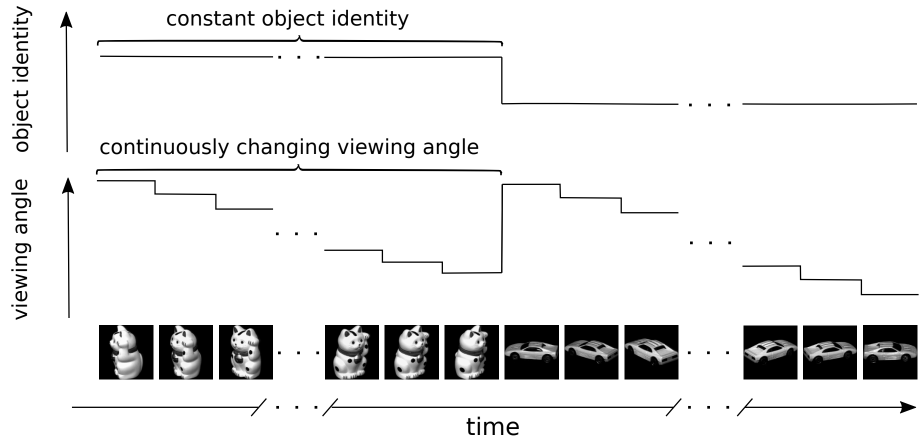


Figure 1.3: Slow and fast changing features. In natural viewing situations, e.g. watching an object in our hands while rotating it, properties related to the viewing angle change fast and continuously, whereas object identity stays constant until we decide to look at a different object.

has shown how temporal proximity can be utilized to learn invariant representations. He proposed a new synaptic learning rule that incorporates a decaying trace of previous cell activity:

> "A learning rule is therefore needed to specify these modifiable simple-to-complex connections. A simple Hebbian rule, which depends only on instantaneous activations, does not work here as it only detects overlapping patterns in the input and picks up correlations between input units." (Földiák, 1991)

Földiák demonstrated in a neural network that uses this *trace rule* for adjusting forward connections, how orientation selective cells emerge that are similar to complex cells in the primary visual cortex (Hubel and Wiesel, 1962). After the network was trained with sequences of moving edges, these cells showed high selectivity for a preferred orientation but responded invariantly to the same orientation at different positions. When applied in a hierarchical network, the trace rule can enable invariant responses to complex stimuli such as hand written characters (Wallis, 1996) or faces (Wallis and Rolls, 1997).

Several mechanisms have been proposed by which something equivalent to the trace rule could be realized in the brain. First, high neural activity could trigger the release of chemicals such as nitric oxide to be used as a signal for learning (Földiák, 1992). Second, binding of glutamate to N-methyl-D-aspartate receptors (NMDAR) for 100 ms or more could provide a cellular basis for the trace rule (Rolls et al., 1992; Földiák, 1992). Third, the trace rule might not be implemented within a single cell. Instead, persistent firing of neurons could enable the association of subsequent images (Rolls and Tovee, 1994). One aim of this dissertation is to explore this third mechanism in a spiking neural network (section 3.1).

Self-Organizing Topographic Maps

In many cortical areas response properties of neurons are mapped continuously along the cortical surface (Kaas, 1997). E.g. a topography for orientation was found in the primary visual cortex (for example Bosking et al., 1997), whereas a topography for stimulus frequency was found in early areas of the auditory cortex (Saenz

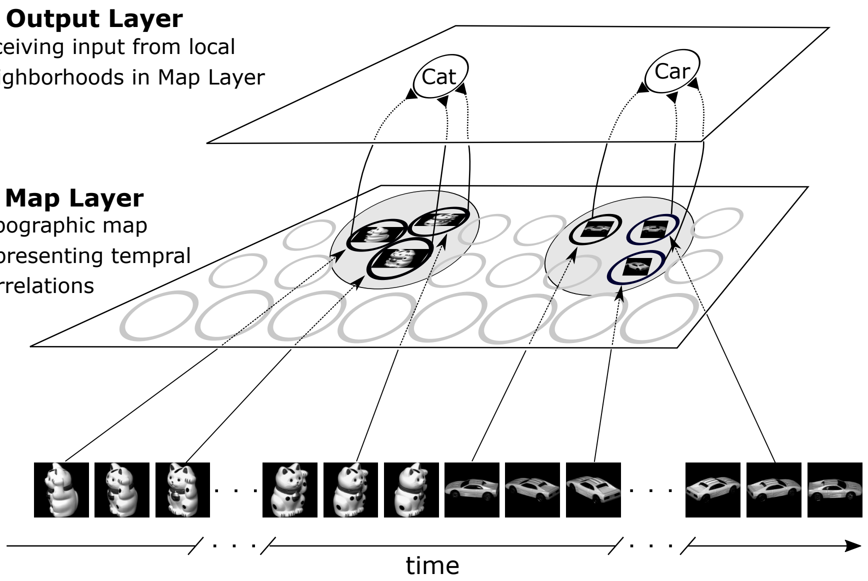


Figure 1.4: Sketch of the invariance mechanism proposed by Michler, Eckhorn, and Wachtler (2009). Different views of the same object are experienced in a sequence. Because of their temporal correlations, views of the same object are represented by neighboring neurons in the *map layer* E1. Neurons in the *output layer* E2 receive input from local neighborhoods in E1. They exhibit invariant responses because of the object topography in E1.

and Langers, 2014; Leaver and Rauschecker, 2016). Experimental data measured in the inferotemporal cortex suggests that higher-order features related to invariant object representations might be mapped in a continuous manner (Wang, Tanaka, and Tanifuji, 1996; Tanaka, 1996; Tanaka, 2003).

Self-Organizing Topographic Maps (SOMs) are a type of neural network models that explain how a topographic order of response properties can emerge based on correlations in their sensory input (Kohonen, 1982; Choe and Miikkulainen, 1998). A SOM network is composed of two dimensional layers of neurons. Each neuron has short range excitatory lateral connections to its neighbors. Competition is introduced by long range lateral inhibitory connections. After training, neighboring neurons show selectivity for similar stimulus patterns. By integrating over a local neighborhood of neurons, a readout mechanism (e.g. a layer of output neurons) can achieve a generalization across sets of similar stimuli.

1.4 Hypotheses and Objectives

The aim of this work is to gain insights into mechanisms underlying visual object recognition in the brain, by simulating the proposed mechanisms in biologically plausible spiking neural networks. Specifically, four hypotheses were investigated. The first three hypotheses are related to invariant object recognition, whereas the fourth is concerned with the discrimination of very similar patterns.

Hypothesis 1 - Sustained Neural Activity can Serve as a Trace Rule

Whereas a lot of biological evidence is available for Hebbian synaptic plasticity (Markram et al., 1997; Bi and Poo, 1998; Dan and Poo, 2006), no evidence for the existence of a synaptic trace rule as proposed by Földiák (1991) has so far been reported in literature. The first hypothesis of this work is that a memory trace for temporal proximity based learning can be provided by the intrinsic dynamics of a network. Rolls and Tovee (1994) have found evidence for sustained firing of cortical neurons for 200-300 ms after presentation of visual stimuli.

Short range excitatory lateral connections could enable continued firing of neurons within the local neighborhood. Once activated, nearby neurons have an increased chance of firing for successive stimuli. Their activity coincides with activity caused by the next stimulus within a sequence, and Hebbian plasticity rules that operate on a short time scale can capture temporal correlations on a longer time scale.

A challenge for this proposed mechanism is the balance between intrinsically generated activity, and activity caused by feedforward connections. When excitatory lateral connections are too strong, intrinsic activity is not be affected by afferent connections, and the network does not learn any representation of presented input patterns. On the other hand, when excitatory lateral connections are too weak, persistent firing can not be sustained, and there is not be a memory trace to associate successive stimuli. Biologically plausible parameters that can influence this balance are the proportion of NMDA and AMPA receptors, synaptic time constants, and synaptic depression (Tsodyks, Pawelzik, and Markram, 1998)

Hypothesis 2 - Topographic Maps can Represent Temporal Correlations

In classical models of self-organizing maps (SOM; section 1.3), the structure of learned maps reflects the statistics of spatial correlations within the set of training stimuli. The second hypothesis is that temporal correlations can be represented in a self-organizing map as well. Because neighboring views of the same object are often seen in a temporal sequence, sustained firing of local groups of neurons can map successive input patterns onto neighboring neurons (Figure 1.4). To separate the effects of spatial and temporal correlations, I created stimulus sets with identical spatial correlations along the axis of a 2D parameter space (named "X-parameter" and "Y-parameter" in Figure 2 on page 26). By training the network with temporal correlations along one axis or the other, differences between learned maps can be attributed to changes in temporal correlations.

Hypothesis 3 - Topographic Maps can Enable Invariance for 3D Rotation

In the *neocognitron* model (Fukushima, 1980), complex cell layers receive input from a local neighborhood within the preceding simple cell layer. Because simple cells of the same layer share the same pattern of synaptic weights, but differ with respect to the corresponding position in the visual field, complex cells achieve translation invariance. If the topographic order of simple cells represents 3D rotation instead, complex cells pooling over neighboring simple cells can exhibit invariant activity with respect to changes of the 3D viewing angle. The invariance of complex cell responses can be tested by measuring their activity for all trained stimuli, and then calculating tuning curves for stimulus parameters like viewing angle and object identity (see equations 15 to 18 and Figure 3 on page 27).

The aim of chapter 3.1 is to develop a proof-of-principle for hypotheses 1 - 3 by combining the concept of temporal proximity based learning with self-organizing topographic maps in a spiking neural network, and testing it by using stimulus sets that allow to separate the effects of temporal and spatial correlations.

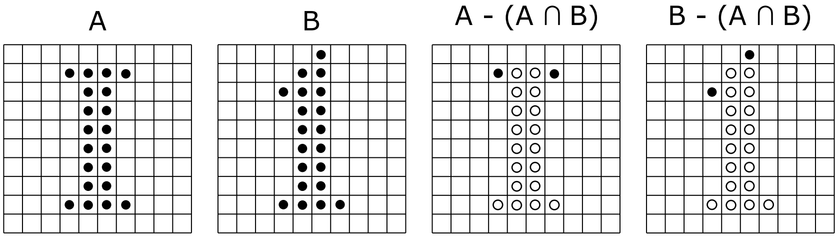


Figure 1.5: Patterns with large overlap. Two patterns A and B with 20 active pixels each, defined in a 10x10 grid. A and B have an overlap ($A \cap B$) of 90 % (only two out of twenty pixels differ). Suppressing the overlapping part of input patterns enhances differences, and can improve discrimination learning.

Hypothesis 4 - Adaptive Feedback Inhibition can Improve Learning

Pattern discrimination is a prerequisite for object recognition. As our own preliminary simulations have shown, a standard approach for pattern discrimination based on Hebbian learning and competition via lateral inhibition can achieve selectivity for stimulus sets with moderate overlap, whereas discrimination performance deteriorates for high overlap (Michler, Wachtler, and Eckhorn, 2006). For very similar patterns, output neurons that respond well to one stimulus also have a high chance of responding well to other stimuli, because they are driven by the overlapping part of input patterns (Figure 1.5). Suppressing that overlap therefore enhances differences and can improve pattern discrimination for very similar stimuli.

My hypothesis is that adaptive inhibitory feedback connections can enable this overlap suppression and therefore improve pattern discrimination. The goal of the publication presented in chapter 3.2 is to provide a proof-of-principle for this hypothesis by implementing it in a network of spiking neurons with STDP based learning rules.

Chapter 2

Methodological Background: Simulating Neural Networks

2.1 Modeling: The Art of Simplification

Mathematical models and computer simulations can help to improve our understanding of complex biological systems. From models, predictions for new experiments can be generated, and proposed ideas about biological mechanisms can be explored to find out whether they actually work as proposed or not. When creating models, many crucial decisions must be made about the level of detail or abstraction. The more biological details a model incorporates, the easier it is to relate the model to the actual biological system. With more detail a model also grows in complexity, which makes it harder to understand how it actually works. Therefore, the goal of modeling is to simplify as much as possible, but keep the essence of what is "important" for the way a biological system solves a problem.

In the last two decades many technical approaches have been developed to tackle object recognition problems, using mathematical methods like *Principal Component Analysis* (Nagaveni and Sreenivasulu Reddy, 2014), *Independent Component Analysis* (Delac, Grgic, and Grgic, 2006), or *Fourier Transformations* (Westheimer, 2001; Ryu, Yang, and Lim, 2018). Such models have greatly improved our understanding of the problem domain. However, to understand how such mechanisms are actually implemented in the brain, we need models that are compatible with our knowledge about its basic building blocks.

2.2 Model Neurons

The main properties of neurons that are relevant for modeling spiking neural networks are the membrane potential, generation of action potentials, and synaptic transmission. When modeling networks with large numbers of neurons, single neuron models must be simplified by distinguishing between critical and non-critical properties.

Point Neurons

In a biological neuron the membrane potential can vary across soma, dendrites and axon. Cable theory (Rall, 1959) can be applied to calculate the spread of currents from dendrite to soma, treating dendrites as cylinders with piecewise constant radius (Figure 2.1 B). If only the membrane potentials at the center of these cylinders are considered, the cable model is discretized and reduced to a compartmental model, which consists of a finite number of membrane patches (Figure 2.1 C). Such

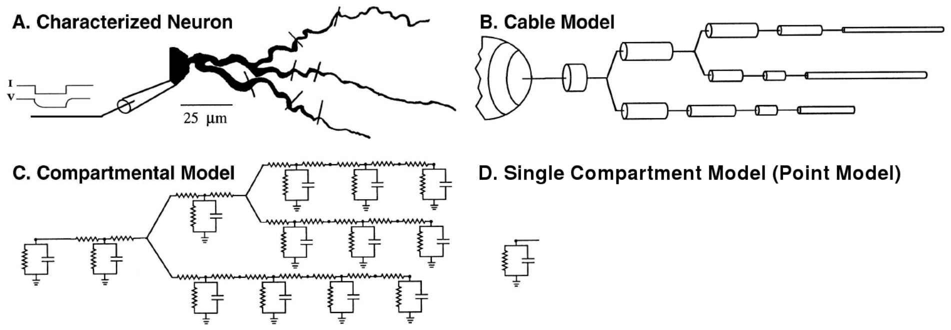


Figure 2.1: Compartmental model vs. point model. Modified from Bower and Beeman (2003). A: Neuron with dendrite and electrodes measuring membrane currents and potentials at the soma and at various positions on dendrites. B: A cable model describes parts of dendrites as cylindric cables in a continuous fashion. C: A compartmental model treats the continuous membrane surface as a finite number of membrane patches. D: In a point model only a single compartment is used.

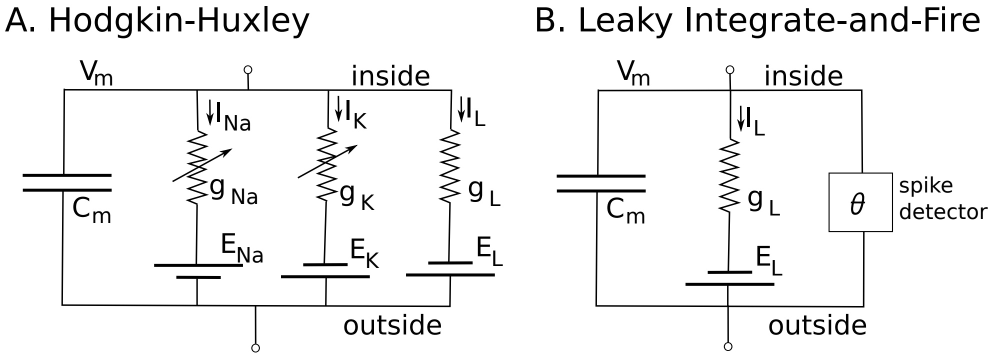


Figure 2.2: Equivalent circuits. A. Equivalent circuit for the Hodgkin-Huxley model. C_m is the capacitance of the lipid membrane. g_{Na} and g_K are voltage dependent conductances for sodium and potassium ions. The *leak conductance* g_L is a constant factor representing all other conductances (mostly for Cl^- ions). The batteries E_{Na}, E_K, E_L represent reverse potentials for respective ion currents. B. Equivalent circuit for the leaky integrate-and-fire-model. It lacks batteries and resistors for voltage dependent sodium and potassium currents. Instead it has a *spike detector* which detects when V_m crosses a threshold θ.

models are used to study interactions between dendrites and the soma. Models that completely ignore the morphology of dendrites and treat the whole neuron as a single compartment are called *point neurons* (Figure 2.1 D). Every incoming input is treated equally, as if every synapse would target the soma. Only a single membrane potential per neuron is calculated. While interactions between dendrites and soma are lost, the drastically reduced computational costs of the point neuron enables simulations with a much larger number of neurons.

Hodgkin-Huxley Neuron

Many neuron models used in neural network simulations are derived from the set of equations formulated by Hodgkin and Huxley in 1952. Figure 2.2 A shows the equivalent circuit for the neuro membrane. The membrane is a capacitor with capacity C_m. Ionic currents are treated as resistors, coupled with a battery according to the equilibrium potential for the respective ions. Since ion channels for sodium (Na^+) and potassium (K^+) are voltage dependent, they are treated as a regulated resistances with conductance g_{Na} and g_K. Currents relying on all other non voltage dependent channels such as Chloride (Cl^-) are summarized as a single *leak current* with conductance g_L. Using voltage clamp experiments with the squid giant axon, Hodgkin and Huxley developed the following set of four differential equations. They describe the dynamics of the membrane potential and the generation of action potentials (APs, often called *spikes*):

$$\begin{aligned} C_m\dot{V} &= -\overbrace{\bar{g}_{Na}m^3h(V-E_{Na})}^{I_{Na}} - \overbrace{\bar{g}_K n^4(V-E_K)}^{I_K} - \overbrace{g_L(V-E_L)}^{I_L} - I_{input} && (2.1)\\ \dot{m} &= \alpha_m(V)(1-m) - \beta_m(V)m && (2.2)\\ \dot{h} &= \alpha_h(V)(1-h) - \beta_h(V)h && (2.3)\\ \dot{n} &= \alpha_n(V)(1-n) - \beta_n(V)n && (2.4) \end{aligned}$$

Differential equations 2.1 to 2.4 describe the dynamics of the membrane potential V in the Hodgkin-Huxley model. C_m is the capacitance of the lipid membrane. $\dot{V} = \frac{dV(t)}{dt}$ is the temporal derivative of V. According to the charging equation for a capacitance $\dot{V} = \frac{I}{C}$, the product $C_m\dot{V}$ is equal to the sum of all currents across the membrane: $I_{Na} + I_K + I_L + I_{input}$, where I_{Na} and I_K are the sodium and potassium ionic currents, I_L the *leak current* and I_{input} any additional input current (e.g. from synaptic currents). The ionic currents depend on the difference of the membrane potential V to their respective reversal potentials E_{Na}, E_K, E_L, and the conductance g for the respective ions. While the leak conductance g_L is a constant, conductances for sodium and potassium are dynamic and voltage dependent. $\bar{g}_{Na}$ and $\bar{g}_K$ are the maximum conductances when all channels are open. m, h, and n are gating variables with values between 0 and 1. They determine the proportion of open sodium and potassium channels $p_{Na} = m^3h$ and $p_K = n^4$. Equations 2.2, 2.3, 2.4 describe the temporal evolution of m, h, and n, depending on their respective voltage dependent variables α and β.

Because the variables in the Hodgkin-Huxley model directly represent biophysical values such as the membrane potential, it is suitable for generating numeric predictions for electrophysiological experiments. About 1200 floating point operations (FLOPS) are needed to simulate the Hodgkin-Huxley model for 1 ms (Izhikevich, 2004). This is computationally expensive. In order to analyze neural network mechanisms that do not rely on the precise values of the membrane potential, simplified models with less computational costs can be used to simulate larger numbers of neurons.

Izhikevich Neuron

Izhikevich (2003) reduced the four dimensional Hodgkin-Huxley equations (2.1) to the following two dimensional system:

$$\begin{aligned} \dot{V} &= 0.04V^2 + fV + e - U + I_{input} && (2.5)\\ \dot{U} &= a(bV - U) && (2.6) \end{aligned}$$

with the auxiliary after-spike resetting:

$$if(V \geq 30mV) \quad then \quad \left\{ \begin{array}{lcl} V & \leftarrow & c \\ U & \leftarrow & U + d \end{array} \right. \qquad (2.7)$$

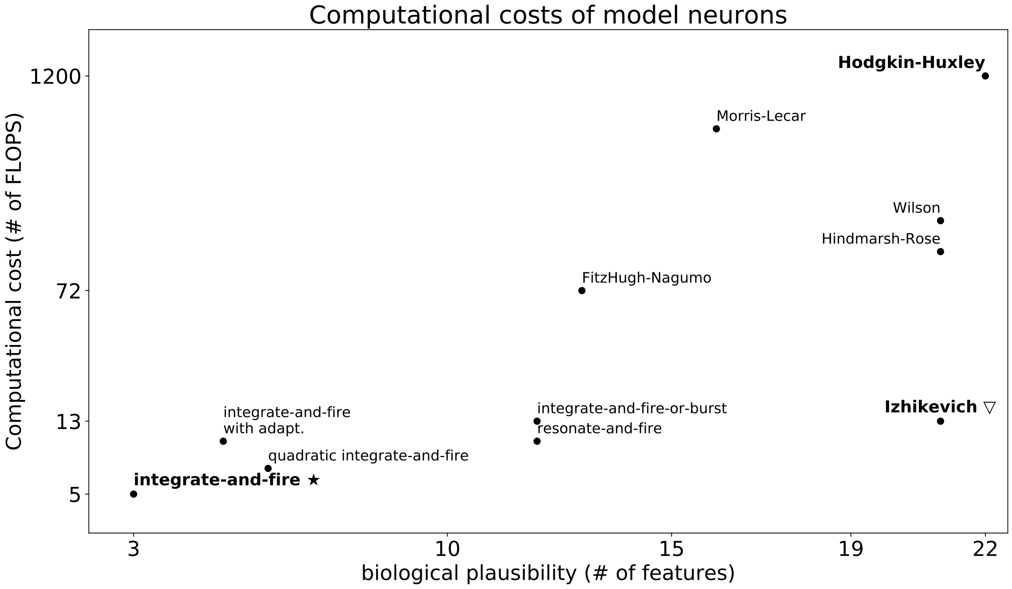


Figure 2.3: Comparison of computational costs and number of neuro-computational features for various model neuron types (modified from Izhikevich, 2004); "# of FLOPS" is an approximate number of floating point operations (addition, multiplication, etc.) needed to simulate the model during a 1 ms time span. "# features" is the number of neuro-computational features as defined by Izhikevich, e.g. the ability of a neuron model to exhibit properties of an *integrator*, or whether it can exhibit burst firing. ★ The *integrate-and-fire* model was used in Michler, Eckhorn, and Wachtler (2009). ▽ The Izhikevich model was used in Michler, Wachtler, and Eckhorn (2006).

V and U are dimensionless variables. V represents the membrane potential. U is a membrane recovery variable, which accounts for the activation of K^+ and inactivation of Na^+ ionic currents. It provides a negative feedback to V. a, b, c, d, e, f are dimensionless parameters. With $f = 5$ and $e = 140$ the spike initiation dynamics of the system approximates the dynamics of a cortical neuron so that the membrane potential V has a mV scale and time t a ms scale.

The reduction to a two dimensional system lowers computational costs down to 13 FLOPS for simulating a neuron for 1 ms, while preserving many dynamic properties of the original Hodgkin-Huxley equations (Figure 2.3). Depending on the choice of parameters, the Izhikevich model can exhibit a variety of excitability patterns. Some examples are:

- tonic spiking: fires continuous train of spikes as long as it is stimulated
- Class 1 excitability: arbitrarily low firing rate, and large range, e.g. 2 - 100 Hz
- Class 2 excitability: no low frequency firing rate; small range, e.g. 100 - 150 Hz
- bursting: many successive spikes with high frequency
- rebound spikes: spikes after inhibitory input
- integrator: successive sub-threshold inputs can cause an AP
- resonator: successive sub-threshold inputs can cause an AP if their delay match the frequency of the intrinsic oscillations.

Izhikevich (2004) describes 20 neuro-computational properties that have been observed in real neurons and can be reproduced with specific parameter values in the Izhikevich model and in the Hodgkin-Huxley model. For the simulations in chapter 3.2 I used model neurons based on Izhikevich's equations.

Leaky Integrate-and-Fire Neuron

A further simplification is the *leaky integrate-and-fire* (LIF) neuron, also known as the *Lapique model* (Lapicque, 1907). As shown in the equivalent circuit in Figure 2.2 only the leak current I_L is considered while omitting the terms for voltage dependent sodium and potassium ion channels.

$$C_m\dot{V} = -\overbrace{g_L(V - E_L)}^{I_L} - I_{input} \tag{2.8}$$

$$if(V \geq V_\theta) \quad then \quad V \leftarrow V_{reset} \tag{2.9}$$

The reverse potential E_L for the leak current I_L is equal to the resting potential. If the membrane potential V temporarily deviates from E_L (due to synaptic input currents I_{input}) it falls back to E_L in an exponential decay.

Due to the missing voltage dependent currents, APs are not generated by internal dynamics of the LIF model. Instead, a threshold V_θ is applied to the membrane potential V. Whenever the threshold is crossed, an AP is generated, and the membrane potential set back to a reset value V_{reset} (equation 2.9). This is depicted as the *spike detector* in Figure 2.2 B.

These simplifications reduce the cost to 5 FLOPS per 1 ms simulation time (see Figure 2.3). The LIF neuron has only 3 of the 20 neuro-computational features listed in Izhikevich (2004): it is Class 1 excitable; it can fire tonic spikes with constant frequency, and it is an integrator. For analyzing mechanisms that do not depend on further features like spike frequency adaptation or bursting, the LIF is a good choice. Because of its low computational cost, large numbers of neurons can be simulated efficiently. Therefore, it was chosen for simulating learning of topographic maps based on spatiotemporal correlations in chapter 3.1 in a network of more than 10.000 neurons.

2.3 Layers

When describing the architecture of an artificial neural network, the term *layer* refers to different levels of the processing hierarchy. Often neural networks have an input layer, one or many processing layers (sometimes referred to as *hidden layers*), and an output layer.

On the implementation level, layers are groups of neurons that share common properties and algorithms. Neurons within a layer typically use the same model type, parameters, and connectivity patterns. Therefore, inhibitory and excitatory neurons are often in separate implementation layers but can represent neurons of the same layer within an anatomical cortex column.

2.4 Synaptic Transmission

Signal transmission between neurons is mediated by electrical and chemical synapses.

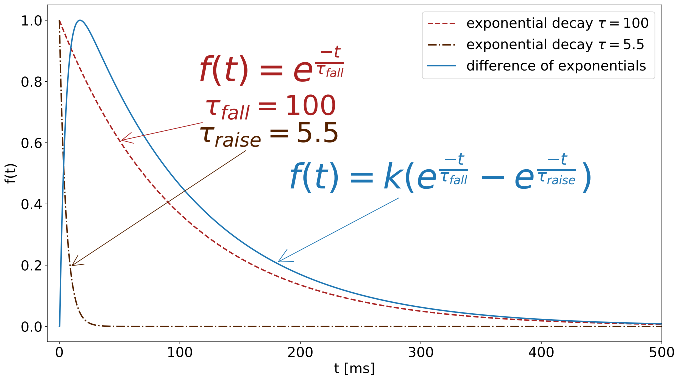


Figure 2.4: Exponential decay and difference of exponentials. The red dashed line shows an example of an exponential decay with $\tau = 100\ ms$ (brown dashed dotted line: $\tau = 5.5\ ms$). Such functions can be used to model processes that fall back to a base line after a deviation; e.g. the amount of transmitter molecules in the synaptic cleft. If the process has a rising phase that can not be neglected, a difference of two exponentials can be used. Blue line: difference of exponentials with $\tau_{fall} = 100\ ms$ and $\tau_{rise} = 5.5\ ms$ used to model NMDA transmitter concentration in Michler, Eckhorn, and Wachtler (2009). K is a constant factor depending on τ_{fall} and τ_{raise} to scale the function so that the peak has a value of 1.0.

Electrical Synapses

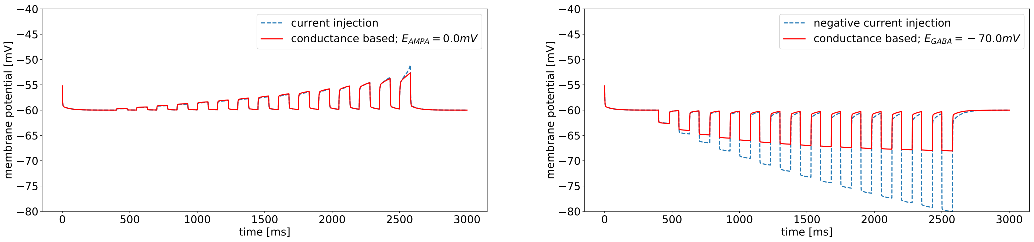
Electrical synapses are fast because currents flow directly between two cells via gap junctions. They can play a role for synchronization, regulation of neural circuits, and retinal feature selectivity (e.g. Nath and Schwartz, 2017). Since they were not used in the studies presented in this dissertation, I will not further discuss them here.

Chemical Synapses

Once an action potential arrives at a chemical synapse, transmitter molecules are released into the synaptic cleft, and ion channels in the postsynaptic membrane are opened, increasing conductance of respective ion currents. Depending on the type of transmitter, this causes an inhibitory or excitatory postsynaptic current (IPSC or EPSC). The amount of active transmitter molecules in the synaptic cleft then decreases. Either they are chemically inactivated (like acetylcholine, which is split into acetate and choline), or they are reabsorbed into the presynaptic membrane by special transporter proteins (like glutamate, GABA, and serotonine; this process is called *reuptake*).

The temporal evolution of the amount of transmitter can be modeled using an exponential decay function. To also consider a raising phase (e.g. the slow activation of NMDA receptors), a difference of exponentials can be used (Figure 2.4).

The simplest way to model postsynaptic currents is to assume they are proportional to the amount of transmitter molecules, and implement it as a current injection (I_{input} in equation 2.8) that is directly added to the membrane potential (like in chapter 3.2 for excitatory synapses). This is a sufficient approximation for excitatory currents, since outside of action potentials the variation of membrane potentials ($-70\ mV$ to $-55\ mV$) is small compared to the difference $V - E_{rev}$ between average membrane potential and reversal potential for excitatory currents ($E_{rev} \approx 0\ mV$ for glutamate receptors).



(a) excitatory **(b)** inhibitory

Figure 2.5: Current injection vs conductance based synaptic input. Membrane potential of an Izhikevich model neuron for a series of rectangular synaptic inputs. For current injection (blue dashed lines) rectangular pulses are directly used as I_{input}. For conductance based input (solid red lines) the difference of reverse potential and membrane potential is considered: $I_{input} = g(V - E_{rev})$. (a) For subthreshold excitatory inputs the difference between current injection (blue dashed line) and conductance based input (red solid line; $E_{AMPA} = 0\ mV$) is very small. (b) For inhibitory inputs, current injection (blue dashed line) lowers the membrane potential with every step, while for conductance based inhibitory input (red solid line; $E_{GABA} = -70\ mV$) the membrane potential converges towards a lower boundary.

Inhibitory Cl^- currents have a reversal potential $E_{rev} \approx -70\ mV$, which is close to the resting membrane potential. Even for very large inhibitory input, the membrane potential would never fall below E_{rev}. Simply adding negative currents would therefore result in unrealistically low membrane potentials (blue dashed line in Figure 2.5b). Conductance based models consider this by calculating the synaptic current from the conductance g_i and the difference between membrane potential and the reverse potential $V - E_{rev}$. Figure 2.5 demonstrates the difference of current injection and conductance based synaptic input for a series of increasing excitatory (Figure 2.5a) and inhibitory (Figure 2.5b) rectangular synaptic inputs.

2.5 Cellular Mechanisms of Neural Plasticity

While the precise mechanisms underlying synaptic plasticity are not yet fully understood, experimental results suggest that for at least one mechanism intracellular Ca^{2+} levels play a crucial role (Shouval, Bear, and Cooper, 2002; Dan and Poo, 2004). Spike timing dependent plasticity (STDP) was found to depend on NMDA receptors (NMDARs) and backpropagating action potentials (Markram et al., 1997; Bi and Poo, 1998).

NMDARs are voltage gated glutamate channels that are permeable for Na^+, K^+, and Ca^{2+}. For membrane potentials near the resting potential (-70 mV) NMDARs stay closed, even if they bind glutamate. This is caused by a Mg^{2+} ion that is part of the receptor and blocks the channel. Once the membrane potential shifts towards less negative values, the position of the Mg^{2+} ion within the NMDAR changes and the channel opens. Because NMDAR activation depends on two factors – transmitter binding and depolarized membrane potential – they can act as coincidence detectors.

When a cell fires an action potential (AP), this AP not only travels along the axon but also propagates back into the cell's own dendrites. There it can interact with NMDARs. Therefore, when the postsynaptic cell fires shortly after the presynaptic cell (pre → post), a backpropagating AP can open NMDARs that have already bound glutamate due to a preceding presynaptic AP. This causes a fast and large

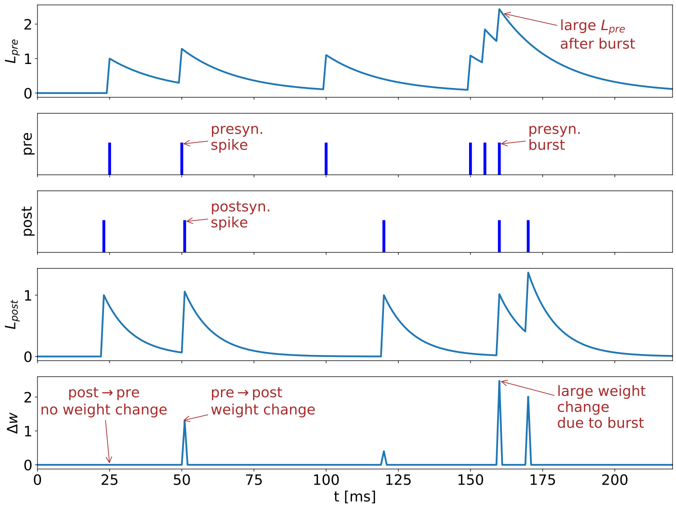


Figure 2.6: Implementation of a Hebbian learning rule. Time course of learning potentials L_{pre}, L_{post}, and weight change Δw for a series of spikes, according to the learning rule used in Michler, Eckhorn, and Wachtler (2009). When a presynaptic spike immediately precedes a postsynaptic spike, both learning potentials L_{pre} and L_{post} are high, and the synaptic weight is increased by Δw (at 50 ms). For a reversed order of presynaptic and postsynaptic spikes (around 25 ms), L_{pre} is still zero at the time Δw is calculated, and therefore the synaptic weight does not change.

increase of Ca^{2+} concentration in the dendrite, which can be used as an intracellular signal to trigger LTP. For the reverse spiking order (post → pre) the backpropagating AP does not coincide with glutamate binding of NMDAR. Raise of Ca^{2+} is therefore small during EPSP, which can be used as a signal to weaken the synapse (LTD).

The time differences between post- and presynaptic spike where significant LTP occurs (critical window) are in a range of 0 - 10 ms (for rat hippocampal slices) and 0 - 40 ms (Xenopus tadpole; review by Dan and Poo, 2006). For LTD the smallest critical windows were 0 to -7 ms (Zebra finch), whereas the largest were 0 to -200 ms (rat hippocampal slice culture).

2.6 Synaptic Learning Rules

The Hebbian learning rule for excitatory synapses used in Michler, Eckhorn, and Wachtler (2009) and Michler, Wachtler, and Eckhorn (2006) is based on *learning potentials* L_{pre} and L_{post} that represent intracellular signals associated with action potentials (e.g. Ca^{2+}concentration and glutamate binding with NMDAR). These variables increase for every presynaptic or postsynaptic spike and then decrease exponentially.

$$w_{n,m} = \delta_m(t_m) R L_{pre,n} L_{post,m} \tag{2.10}$$

$$L_{pre,n} = \sum_{t_n} e^{-\frac{t-t_n}{\tau_{pre}}} \tag{2.11}$$

$$L_{post,m} = \sum_{t_m} e^{-\frac{t-t_m}{\tau_{post}}} \tag{2.12}$$

Synaptic weights are updated with every postsynaptic spike. Mathematically this is expressed by multiplying with a Dirac function $\delta_m(t_m)$ that is 1 at time t_m of a postsynaptic spike, and 0 otherwise. R is a constant to adjust the learning rate. Figure 2.6 shows an example for a series of pre- and postsynaptic spikes.

2.7 Competition: The Winner Takes it All

Competition Between Neurons

Neurons can compete with each other for activation via lateral inhibition (Figure 1.1). The neuron that receives the strongest input suppresses activity of its competitors by activating inhibitory interneurons. Because synaptic plasticity depends on spike frequency, the most active neurons adjust their weights to match the current input pattern. By reducing the number of spikes of competing neurons, the "winner" prevents other neurons from learning the same pattern. The connection between lateral inhibition and learning was already proposed by Grossberg (1969). In the context of computational models of neural networks this principle is known as *winner-take-all* (WTA).

Competition Between Synapses

Hebbian plasticity increases synaptic weights based on correlation between pre- and postsynaptic activity. This creates a positive feedback loop, because increased weights in turn increase correlations. If synaptic weights were allowed to grow unconstrained, the neural network could run into a dysfunctional state with too much activity where no useful information processing takes place anymore (e.g. like an epileptic seizure). To solve this stability problem, synaptic normalization rules can be used that keep the total sum of synaptic strength converging onto one cell constant (von der Malsburg, 1973): as one synapse grows stronger, others are weakened, creating competition between synapses targeting the same neuron.

The underlying cellular processes could be competition for limited resources like dendrite building material and receptor molecules, or a form of spike timing dependent synaptic depression that balances the total amount of synaptic input. Further, a variety of homeostatic plasticity phenomena have been found (Turrigiano and Nelson, 2004). Modeling results by Zenke, Hennequin, and Gerstner (2013) suggest the existence of a homeostatic regulatory mechanism that reacts to firing rate changes on the order of seconds to minutes.

Chapter 3

Publications

3.1 Using Spatiotemporal Correlations to Learn Topographic Maps for Invariant Object Recognition

Summary

In the following publication "Using Spatiotemporal Correlations to Learn Topographic Maps for Invariant Object Recognition" (Michler, Eckhorn, and Wachtler, 2009) we address the problem of invariant object recognition in spiking neural networks. We propose a new mechanism that combines two established principles of neural computation in a novel way to enable unsupervised learning of viewpoint invariant representations of visual objects: learning based on temporal contiguity and the formation of self-organizing topographic maps (SOMs). Our main hypotheses are:

1. Temporal correlations in input sequences can shape the neighborhood relations in a topographic map.
2. A feature topography that reflects spatial **and** temporal correlations can support viewpoint invariant coding of object identity.
3. Intrinsically sustained spiking activity can provide a memory trace suitable to bind successively observed views of objects to representation that enables invariant recognition.

We used stimuli that allowed us to separate the effects of spatial and temporal correlations. By changing the order of stimuli during learning we show that the differences of learned topographic maps indeed reflect temporal correlations.

Our results show that in spiking neural networks learning based on temporal contiguity is possible without the need of a new mechanism of spike timing dependent synaptic plasticity (STDP) that operates on a longer time scale. Instead, lateral connections between excitatory neurons can sustain the spiking activity of a local group of neurons, thereby providing a memory trace with a functional role similar to a synaptic *trace rule*. Our model suggests that the topographic order of feature representations observed in various parts of the visual cortex has a functional role for invariant object recognition.

Declaration of Own Contributions

- All simulations presented in this dissertation were implemented by myself using a C++ based **obj**ect-oriented **sim**ulation library (OBJSIM) for spiking neural networks, which I developed. The source code repository is now published and available along with contributions by Dr. Sebastian Thomas Philipp at https://gin.g-node.org/FrankMichler/ObjSim (Michler and Philipp, 2020).

- I developed graphical user interfaces to setup network simulations, visualize network activity and simulation results, and adjust network parameters.

- I implemented the network architecture for learning topographic maps using OBJSIM.

- I designed stimulus sets that are arranged in a 2D feature space to separate effects of spatial and temporal correlations.

- I created 3D models and animations with rotating objects using *Crystal Space*, an open source 3D rendering engine (Tyberghein et al., 2007), to be used as realistic but controllable network input.

- I conducted simulations and parameter scans on the computing cluster *MaRC* of the University Computer-Center of Philipps-University Marburg.

- I wrote the manuscript in collaboration (discussions, suggestions, editing) with Prof. Dr. Thomas Wachtler and Prof. Dr. Reinhard Eckhorn.

- The article “Using Spatiotemporal Correlations to Learn Topographic Maps for Invariant Object Recognition” was peer reviewed by three anonymous reviewers and published as presented here in *Journal of Neurophysiology* (Michler, Eckhorn, and Wachtler, 2009).

J Neurophysiol 102: 953–964, 2009.
First published June 3, 2009; doi:10.1152/jn.90651.2008.

Using Spatiotemporal Correlations to Learn Topographic Maps for Invariant Object Recognition

Frank Michler, Reinhard Eckhorn, and Thomas Wachtler
NeuroPhysics Group, Philipps-University Marburg, Marburg, Germany

Submitted 9 June 2008; accepted in final form 29 May 2009

Michler F, Eckhorn R, Wachtler T. Using spatiotemporal correlations to learn topographic maps for invariant object recognition. *J Neurophysiol* 102: 953–964, 2009. First published June 3, 2009; doi:10.1152/jn.90651.2008. The retinal image of visual objects can vary drastically with changes of viewing angle. Nevertheless, our visual system is capable of recognizing objects fairly invariant of viewing angle. Under natural viewing conditions, different views of the same object tend to occur in temporal proximity, thereby generating temporal correlations in the sequence of retinal images. Such spatial and temporal stimulus correlations can be exploited for learning invariant representations. We propose a biologically plausible mechanism that implements this learning strategy using the principle of self-organizing maps. We developed a network of spiking neurons that uses spatiotemporal correlations in the inputs to map different views of objects onto a topographic representation. After learning, different views of the same object are represented in a connected neighborhood of neurons. Model neurons of a higher processing area that receive unspecific input from a local neighborhood in the map show view-invariant selectivities for visual objects. The findings suggest a functional relevance of cortical topographic maps.

INTRODUCTION

Invariant object recognition

Our visual system has the capability of invariant object recognition: we recognize a familiar object under different viewing conditions, despite drastic variations in the corresponding retinal images with viewing angle, distance, or illumination. Physiological studies have shown that cells in monkey V4 and inferotemporal cortex (Ito et al. 1995; Tanaka 1996, 2003; Tovee et al. 1994; Wang et al. 1996) and in the human hippocampus (Quian Quiroga et al. 2005) show selectivity for objects invariant of size or viewing angle.

A prototype for models of invariant representations is the *pooling model* (Hubel and Wiesel 1962; Kupper and Eckhorn 2002; Riesenhuber and Poggio 1999). An output cell receives input from a pool of cells that have the same selectivity in one feature dimension, but a different selectivity in a second feature dimension. The output cell will then respond selectively to the first feature, but will show invariant responses with respect to the second feature.

Spatial and temporal stimulus correlations as cues for learning invariant representations

When we move through our environment while fixating an object, or when we manipulate an object, different views of the same object appear in temporal sequence. The retinal projections change continuously, whereas the identity of the object remains the same. Under such natural viewing conditions, projections of different views of the same object are spatially and temporally correlated. Physiological (Miyashita 1993; Stryker 1991) and psychophysical (Wallis and Bülthoff 2001) studies have shown that these correlations influence the learning of object representations.

Several mechanisms have been proposed for how these correlations could be used for learning invariant representations (Becker 1993; Einhäuser et al. 2002; Földiák 1991; Stringer et al. 2006; Wallis 1996; Wiskott and Sejnowski 2002). Földiák (1991) proposed a modified Hebbian learning rule—the *trace rule*—that exploits temporal correlations in a sequence of input patterns. The trace learning rule has been used in a hierarchical multilayer network, to achieve invariant response properties for more realistic stimuli (Rolls and Stringer 2006; Stringer and Rolls 2002; Wallis and Rolls 1997).

How the trace rule is implemented in cortical circuits is still an open question. Wallis and Rolls (1997) argued that persistent firing, the binding period of glutamate in the *N*-methyl-D-aspartate (NMDA) channels, or postsynaptically released chemicals such as nitric oxide might be the biological basis for the trace rule. Sprekeler et al. (2007) showed theoretically that the learning rule for slow feature analysis (SFA), which is related to trace learning, can be achieved with spiking neurons. Nevertheless, invariance learning on the basis of temporal correlations has not yet been implemented in a network of spiking neurons.

Previous models for invariance learning (Einhäuser et al. 2002; Riesenhuber and Poggio 1999; Wallis and Rolls 1997) relied on not only the learning of features but also learning the specific connections to pool across related features to achieve invariant representations. We will show that feature representations can be learned in an ordered way, such that related features are represented in a local neighborhood and invariance can be achieved by a generic connectivity without the need for further learning. The key mechanism for this is to learn a topographic map that reflects the spatiotemporal correlations of the inputs.

Topographic maps and spatiotemporal stimulus correlations

Many cortical areas are topographically organized. In primary visual cortex (V1), neighboring neurons receive input from neighboring parts of the retinal image. Superimposed on the retinotopic organization is an orientation topography: neighboring populations of neurons respond to edges of similar orientation (Hubel and Wiesel 1974). In inferotemporal cortex,

---

Address for reprint requests and other correspondence: F. Michler, NeuroPhysics Group, Philipps-University Marburg, Renthof 7, 35032 Marburg, Germany (E-mail: Frank.Michler@physik.uni-marburg.de).

0022-3077/09 $8.00 Copyright © 2009 The American Physiological Society

Downloaded from www.physiology.org/journal/jn (084.134.067.044) on May 31, 2019.

topography for more complex features or even for characteristics of object views was found (Wang et al. 1996). This suggests that some higher-order features of the input are mapped continuously in a topographic fashion (for review see Tanaka 1996, 2003).

The model for the self-organization of cortical maps proposed by von der Malsburg (1973) relies on Hebbian learning in forward connections, short-range lateral excitation, and long-range lateral inhibition. A biologically realistic implementation of this learning principle is the RF-SLISSOM (receptive field–spiking laterally interconnected synergetically self-organizing map; Choe and Miikkulainen 1998) model, which uses spiking model neurons. Trained with a stimulus set of oriented bars, these models can learn orientation maps similar to those found in primary visual cortex. In these studies, stimuli were presented in pseudorandom order to exclude the effects of temporal correlations. As our results show, temporal correlations can affect the emerging topography in this model architecture, if the lateral connections have a large time constant.

An attempt to extend the von der Malsburg model to account for temporal correlations has been considered by Wiemer and colleagues (Wiemer 2003; Wiemer et al. 2000). It is based on lateral propagation of activity, but has not been implemented in a biologically realistic network.

Goals and hypotheses

In this study we investigate a learning principle that combines the idea of spatial and temporal correlation-based invariance learning with self-organizing map formation. Hebbian learning suggests that the emerging topography of a self-organizing network with slow lateral connections is influenced not only by spatial but also by temporal correlations (Saam and Eckhorn 2000). In this study our main hypothesis is that temporal correlations in input sequences can shape the neighborhood relations in a learned topographic map. Furthermore, we hypothesize that a feature topography that reflects spatial and temporal correlations can support the view-invariant coding of object identity. We investigated these hypotheses with simulations of a biologically plausible network of spiking neurons. The slowness principle for learning invariant representations can be implemented in a biologically realistic spiking neural network by using NMDA-mediated short-range lateral connections and long-range lateral inhibition. This connectivity can cause a network dynamics with persistent activity that implements a memory trace. By manipulating the temporal correlations of the input we systematically investigated the effects of stimulus similarity and temporal proximity. View invariance is achieved by neurons of a downstream area that receive input from the topographic map via fixed, generic connections.

METHODS

Network architecture

The network consists of a forward pathway of three layers of spiking neurons. Layer E0 is the input layer, layer E1 represents the map formation layer, and the output layer E2 represents a cortical stage downstream of layer E1 (Fig. 1). Neurons in layers E0 (30 × 30 or 8 × 24 × 26 neurons, depending on the stimulus set), E1 (100 × 100), and E2 (10 × 10) are arranged in two-dimensional (2D) arrays. E0 neurons are activated by the stimulus patterns (see following text). E0 has α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA)–mediated excitatory forward projections ($W_{E1,E0}$) to the excitatory neurons of layer E1. These connections exhibit Hebbian plasticity. The connectivity from E0 to E1 is initially all-to-all with equal weights.

In addition to input from E0, E1 neurons receive excitatory input from their neighbors ($W_{E1,E1}$) with fixed connection strengths that decrease with the distance between two neurons according to a Gaussian

$$w(E1,\,E1)_{i,j} = \begin{cases} S_{E1,E1}\ \exp\left[-\dfrac{1}{2}\left(\dfrac{d_{i,j}}{\sigma_{E1,E1}}\right)^{2}\right] & i \neq j \\ 0 & i = j \end{cases} \qquad (1)$$

where $w(E1, E1)_{i,j}$ is the synaptic strength (weight) of the connection from neuron j to neuron i, $S_{E1,E1}$ is the maximum connection strength, $d_{i,j}$ is the Euclidean distance between neurons j and i, and $\sigma_{E1,E1}$ is the width of the Gaussian kernel. We used toroidal boundary conditions to avoid boundary effects. E1 neurons mutually inhibit each other via a pool of inhibitory interneurons I1. The connectivity between E1 and I1 is random; thus the pool of inhibitory neurons (I1) has no topographic order. Lateral excitatory connections from E1 to E1 and from E1 to I1 are mediated via fast AMPA (τ_{decay} = 2.4 ms) and slow NMDA (τ_{decay} = 100 ms) currents. Inhibitory connections from I1 to I1 and I1 to E1 are mediated by a γ-aminobutyric acid type A

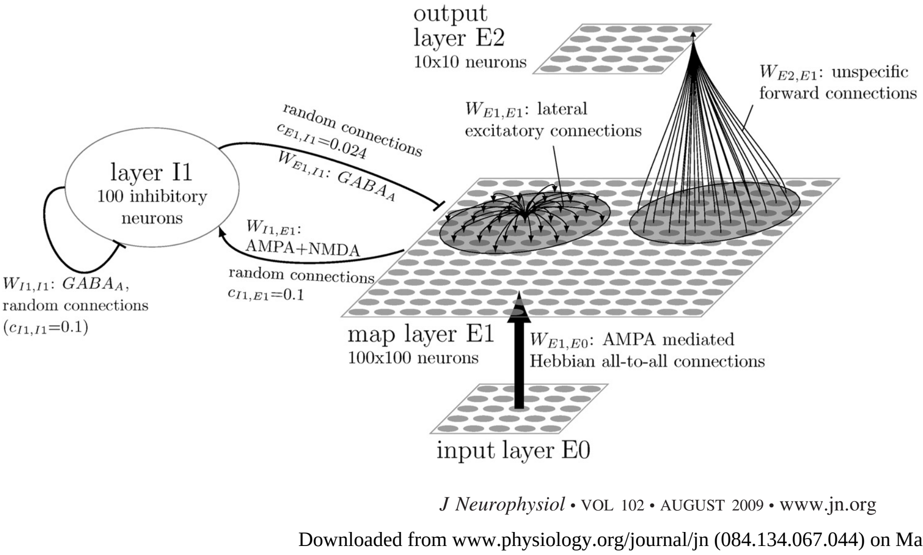


FIG. 1. Model architecture. The model consists of 3 layers of excitatory neurons (E0, E1, E2). Hebbian forward connections from E0 to E1 are all-to-all ($W_{E1,E0}$). Lateral excitatory connections ($W_{E1,E1}$) between E1 neurons are restricted within a lateral interaction range. Each E1 neuron has connections ($W_{I1,E1}$) to a random subset of the inhibitory interneurons I1. I1 neurons have inhibitory connections ($W_{E1,I1}$) to a random subset of I1 and E1. Each E2 neuron receives input from a local subregion of E1 ($W_{E2,E1}$).

Downloaded from www.physiology.org/journal/jn (084.134.067.044) on May 31, 2019.

($GABA_A$) current (fast, $\tau_{decay} = 7.0$ ms). E1 neurons project to output layer E2 with a Gaussian weight profile

$$w(E2,\,E1)_{i,j} = S_{E2,E1}\exp\left[-\frac{1}{2}\left(\frac{d_{i,j}}{\sigma_{E2,E1}}\right)^2\right] \tag{2}$$

These connections were fixed and did not change during the simulation. Thus a neuron in layer E2 receives input from a fixed, localized region of layer E1. The connectivity patterns are summarized in Table 1.

Model neurons

Spiking neurons were simulated by a standard leaky integrate-and-fire model with a voltage threshold and biologically realistic synaptic potentials (Brunel and Wang 2001; Deco and Rolls 2005)

$$C_m\frac{dV(t)}{dt} = -g_L[V(t) - E_L] - I_{syn}(t) \tag{3}$$

where C_m is the membrane capacitance and g_L is the leak conductance of the membrane. When the membrane potential exceeds the firing threshold Θ, an action potential (spike) is generated. The downstroke of the spike is modeled by resetting the membrane potential to V_{reset}. After each spike an absolute refractory period of 1 ms duration is introduced. Parameter values are given in Table 2.

Excitatory forward connections are mediated by AMPA currents, lateral excitatory connections are mediated by AMPA and NMDA currents, and inhibition is mediated by fast $GABA_A$ currents. $I_{syn}(t)$ is the sum of the AMPA, NMDA, and $GABA_A$ synaptic currents

$$I_{syn}(t) = I_{AMPA}(t) + I_{NMDA}(t) + I_{GABA_A}(t) \tag{4}$$

$$I_{AMPA}(t) = G_{AMPA}(t)\hat{G}_{AMPA}[V(t) - E_{AMPA}] \tag{5}$$

$$I_{GABA_A}(t) = G_{GABA_A}(t)\hat{G}_{GABA_A}[V(t) - E_{GABA_A}] \tag{6}$$

$$I_{NMDA}(t) = \frac{G_{NMDA}(t)\hat{G}_{NMDA}[V(t) - E_{NMDA}]}{1 + [Mg^{2+}]\exp[-0.062V(t)]/3.57} \tag{7}$$

where $E_{AMPA} = 0$ mV, $E_{NMDA} = 0$ mV, and $E_{GABA_A} = -70$ mV are the reverse potentials for the synaptic currents. The nonlinear voltage dependence of the NMDA current (caused by the Mg^{2+}-block, *Eq. 7*) was modeled according to Jahr and Stevens (1990).

$\hat{G}_{AMPA}$, $\hat{G}_{GABA_A}$, and $\hat{G}_{NMDA}$ are the maximum synaptic conductivities when all channels are open. $G_{AMPA}(t)$, $G_{GABA_A}(t)$, and $G_{NMDA}(t)$ are the respective fractions of open channels. When a presynaptic spike occurs at $t = t_{sp}$, the fraction of open channels $G(t)$ increases and then decreases. This process is modeled with a difference of two exponentials (*Eq. 8*)

$$G(t) = G_{rise}(t) - G_{decay}(t) \tag{8}$$

$$\frac{d}{dt}G_{rise}(t) = -\frac{G_{rise}(t)}{\tau_{rise}} + w_{m,n}e_{m,n}(t)\delta(t - t_{sp}) \tag{9}$$

TABLE 1. *Connection properties*

Connection	Connectivity Schema	Postsynaptic Currents
E0 → E1	All-to-all, modifiable	AMPA (fast)
E1 → E1	Gaussian kernel with range σ_{E1E1}	AMPA (fast) + NMDA (slow)
E1 → I1	Random, connectivity = c_{I1E1}	AMPA (fast) + NMDA (slow)
I1 → E1	Random, connectivity = c_{E1I1}	$GABA_A$ (fast)
I1 → I1	Random, connectivity = c_{I1I1}	$GABA_A$ (fast)
E1 → E2	Gaussian kernel with range σ_{E2E1}	AMPA (fast)

Connectivity and postsynaptic currents are shown for all synaptic connections between neuron layers. Connections between E0 and E1 are modifiable (see text), whereas all other connections are fixed.

TABLE 2. *Model neuron parameters*

Parameter	Excitatory Neurons	Inhibitory Neurons
C_m	0.5 nF	0.2 nF
g_L	25 nS	20 nS
Θ	−50 mV	−50 mV
V_{reset}	−55 mV	−55 mV

Parameters for inhibitory and excitatory neurons were taken from Deco and Rolls (2005).

$$\frac{d}{dt}G_{decay}(t) = -\frac{G_{decay}(t)}{\tau_{decay}} + w_{m,n}e_{m,n}(t)\delta(t - t_{sp}) \tag{10}$$

where $w_{m,n}$ is the synaptic weight and $e_{m,n}$ is the synaptic efficacy. The forward connections from E0 to E1 are not depressive [$e_{m,n}(t) =$ const $= 1$] and evoke only an AMPA current. The recurrent connections between E1 neurons evoke both AMPA and NMDA currents. The ratio between the peak amplitude of NMDA and AMPA currents was set to 0.3 (Crair and Malenka 1995). These recurrent connections show synaptic depression to stabilize the network activity. For the synaptic dynamics we used a simplified version of the model proposed by Tsodyks et al. (1998)

$$\frac{e_{m,n}(t)}{dt} = \frac{1 - e_{m,n}(t)}{\tau_{rec}} - U_{se}e_{m,n}(t_{sp})\delta(t - t_{sp}) \tag{11}$$

where U_{se} is the fraction of available transmitter that is released during a postsynaptic spike and τ_{rec} is the recovery time constant for the transmitter pool.

Learning rule

We used a Hebbian learning rule similar to that proposed by Gerstner et al. (1996), Saam and Eckhorn (2000), and Michler et al. (2006). The synaptic weights $w_{m,}n$ of the forward connections from layer E0 to E1 are adapted according to the following equations

$$\frac{d}{dt}w_{m,n} = \delta_m(t)RL_{pre,n}L_{post,m} \tag{12}$$

$$L_{pre,n} = \sum_{t_{sn}}\exp\left(-\frac{t - t_{sn}}{\tau_{pre}}\right) \tag{13}$$

$$L_{post,m} = \sum_{t_{sm}}\exp\left(-\frac{t - t_{sm}}{\tau_{post}}\right) \tag{14}$$

where $\delta_m(t)$ is 1 when a spike occurs in the postsynaptic neuron m; otherwise, $\delta_m(t)$ is zero. t_{sn} and t_{sm} denote the times of the past pre- and postsynaptic spikes. When a spike occurs, the pre- or postsynaptic learning potentials $L_{pre,n}$ or $L_{post,m}$ are increased by 1. They exponentially decrease with time constants $\tau_{pre} = 20$ ms and $\tau_{post} = 10$ ms. The exact values of these parameters are not critical. R corresponds to the learning rate. Because learning occurs only after postsynaptic spikes [$\delta_m(t) = 1$], this learning rule is temporally asymmetric; it prefers presynaptic before postsynaptic spiking. The learning rule increases weights if pre- and postsynaptic neurons have overlapping spike trains on a short timescale on the order of τ_{pre} and τ_{post}.

Each time the firing rate of a postsynaptic neuron exceeds a threshold ($\theta_{norm} = 50$ Hz), all input weights are multiplied by normalization factor $f_{norm} < 1$. Evidence for normalization of synaptic weights exists (e.g., Royer and Paré 2003), but the mechanisms are not yet understood. Weight normalization prevents infinite growth of weights and introduces competition between the inputs of a neuron.

Downloaded from www.physiology.org/journal/jn (084.134.067.044) on May 31, 2019.

Stimuli

The network was trained with sets of parameterized stimuli that differed along two parameter dimensions, denoted X and Y, respectively. We tested three increasingly complex stimulus sets, with different correlation structures.

GAUSSIAN STIMULI. Gaussian stimuli consisted of 2D Gaussian activity profiles varying in the horizontal and vertical positions of the center of the Gaussian. These coordinates were used as X and Y dimensions of the stimulus space. The correlation structure of this stimulus set is symmetrical in X and Y. Because of this symmetry we can isolate the effects of the temporal correlations by using stimulus sequences with temporal correlations along either the X or the Y direction of the stimulus space.

PRISM STIMULI. We generated a set of stimuli with variation corresponding to viewing angle (X parameter) and object identity (Y parameter) of three-dimensional (3D) objects. Objects were triangular prisms (Fig. 2*A*).

We varied an arbitrary set of parameters of the prism: the height, the size of the top and bottom triangles, the rotation angle between the top and the bottom triangles, and 3D orientation of the top and bottom triangles. Each of these parameters was systematically changed in steps according to a periodic triangular function tri $(Y + \phi)$ (Fig. 2*B*), which maps the parameter values to the Y dimension of the 2D stimulus parameter space. Therefore the shape changed only along a one-dimensional manifold. Shifting the phase of the triangular function ϕ for different parameters, we obtained toroidal boundary conditions for the stimulus deformation parameter Y. An irregular texture was applied to the surfaces of the prisms to make the faces of the prism more distinct (Fig. 2*C*). Using the open source 3D library Crystal Space (Tyberghein et al. 2007) we generated views of these objects, rotated around their vertical axis with a step size of 18°, resulting in a set of 20 × 20 stimulus pictures, each with 200 × 200 pixels (Fig. 2*D*). Stimuli were preprocessed by a set of 30 × 30 Gabor filters of 19-pixel wavelength and 6-pixel width of Gaussian, comprising eight orientations. To reduce the number of input neurons required, the resolution of the input array was reduced to 30 × 30 by resampling and cropped to 26 × 24 pixels. The outputs of these orientation filters were then used as input signals for the E0 neurons.

COIL STIMULI. To test the performance of the network for more natural stimuli we used images of natural objects taken under different viewing angles. Images were taken from the Columbia Object Image Library (COIL-100) database (Nene et al. 1996). We created a stimulus set with 10 objects and 36 views of each object. The X dimensions corresponded to the viewing angle and the Y dimension to object identity. With respect to the prism stimulus set, the pictures were preprocessed by 30 × 30 Gabor filters (eight orientations; 10-pixel wavelength; 2.1-pixel width of Gaussian), resampled to 30 × 30 pixels, and cropped to 26 × 24 pixels. In contrast to the Gaussian and prism stimuli, in this stimulus set there was no continuous transformation along the Y dimension (object identity) of the stimulus space.

Training and test procedures

We used three training conditions with different temporal correlations between the elements of the stimulus set. In the X slow condition the X parameter was held temporally constant for intervals of $t_{const} \in$ [400 ms, 600 ms], whereas the Y parameter was changed continuously. After each of these training intervals, a short interstimulus interval (20 ms) occurred and X and Y parameters switched to random values for the next training interval (see Supplemental Fig. S1).[1] In the Y slow condition temporal correlations were conversed: the Y parameter was held constant for durations of t_{const}, whereas the X parameter changed continuously. Thus temporal correlations were restricted to the fast changing dimension of the stimulus set.

As a control we simulated a random training condition with random order of stimuli in the sequence, i.e., without temporal correlations.

Network simulations were performed with 125-s training epochs in alternation with test epochs. Both training and testing were done with the full stimulus sets. With 20 training epochs for the Gaussian and prism stimuli and 10 training epochs for the COIL stimuli, total simulated training times were 2,500 and 1,250 s, respectively. During the training epochs the forward connections from layer E0 to E1 were adapted according to the Hebbian learning rule.

During test epochs we tested the network properties with the complete stimulus set. Hebbian plasticity was turned off. Each stim-

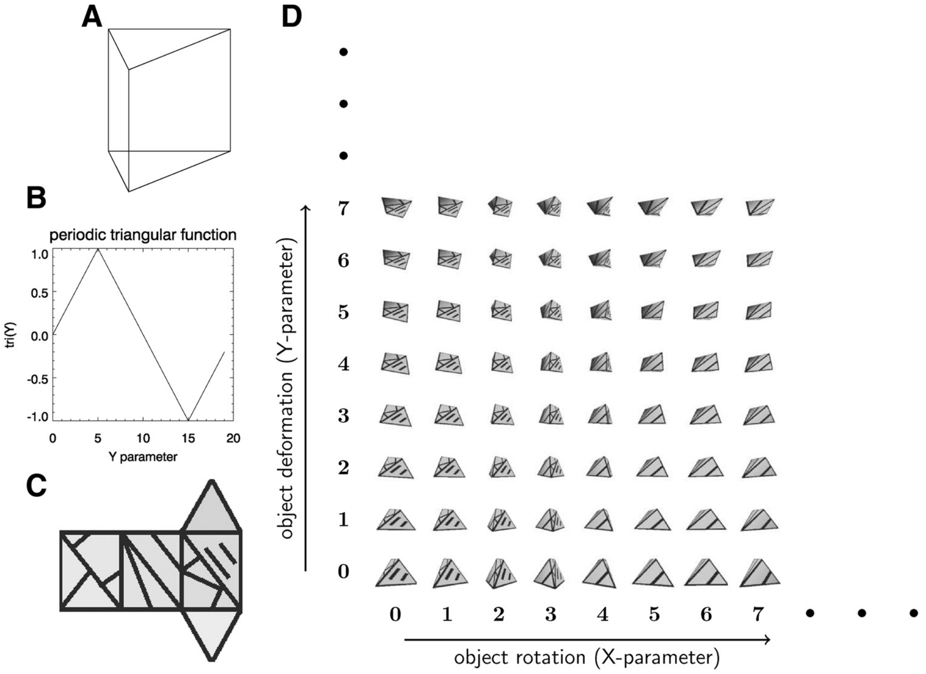

A
D
B
periodic triangular function
1.0
0.5
0.0
-0.5
-1.0
tri(Y)
0
5
10
15
20
Y parameter
C
object deformation (Y-parameter)
0
1
2
3
4
5
6
7
object rotation (X-parameter)

FIG. 2. Three-dimensional (3D) stimulus set. *A*: triangular prism. *B*: periodic triangular function used to continuously change the 3D object parameters along the Y-axis of the stimulus space. *C*: surface texture of the prism. *D*: a 3D-object stimulus set was generated by deforming and rotating the prism (see text).

---

[1] The online version of this article contains supplemental data.

Downloaded from www.physiology.org/journal/jn (084.134.067.044) on May 31, 2019.

ulus was presented for 250 ms. In contrast to learning epochs, in test epochs after each stimulus presentation, all dynamic network variables such as the membrane potentials and synaptic depression parameters were reset to avoid persistent activity evoked by the previous stimulus.

To evaluate how well the stimulus patterns were encoded in the activity of E2 neurons, we determined mean estimation errors for the X and Y parameters. The estimation error measures how reliably information about the currently present stimulus can be read out from the network activity. Because we tested the current network activity with the representation after the penultimate learning epoch, the estimation error is also a measure of stability of the representation. The following equations explain the X estimation error e_x. For the Y estimation error e_y X and Y in the following equations are exchanged.

The X estimation error e_x is the difference between the actual X value of the test stimulus X_n of the last (nth) test epoch and the X value X_p estimated by the network activity based on the tuning curves drawn from the penultimate test epoch. $N_x = 20$ is the size of the X dimension of the stimulus space. The stimulus space is circular (e.g., distance between stimuli 0 and 19 is 1). The difference between two stimuli in this stimulus space is the shortest distance along a circular path

$$e_x = \min\left(\{|X_p - X_n|, \quad |X_p + N_x - X_n|, \quad |X_p - N_x - X_n|\}\right) \quad (15)$$

The estimated X value X_p is calculated by taking into account the activity of all E2 neurons ($a_n[j]$, $j \in \{1, \ldots, J_{E2}\}$), elicited by the current stimulus, and the corresponding tuning curves $T[X, j]$ of the E2 neurons. For a given value X the neural activity of a single neuron j multiplied by the corresponding value of the tuning curve ($T[X, j]$) is a measure for how strong this neuron estimates value X. The sum of this measure over all neurons is the population prediction $P[X]$

$$P[X] = \sum_{i=1}^{J_{E2}} (T[X, j]a_n[j]) \quad (16)$$

The estimated value X_p is the one with the highest likelihood

$$X_p = \arg\max(P[X]) \quad (17)$$

The tuning curve $T[X, j]$ is calculated using the E2 responses of the penultimate test epoch ($n - 1$)

$$T[X, j] = \frac{1}{N_y} \sum_{Y=0}^{N_y-1} a_{n-1}[X, Y, j] \quad (18)$$

The original preference indices are in the range from 0 to 19. Because of the toroidal boundary conditions values 0 and 19 are direct neighbors in stimulus space. Therefore the maximal difference is 10. Note that for a uniform distribution, estimation error values of 0 and 10 would have a probability of 5%, whereas because of the rectification (*Eq. 15*), the values 1 to 9 would have a probability of 10%.

For a representation that is invariant with respect to the X parameter and selective for the Y parameter, the mean estimation error for the Y parameter e_y would be low and the mean estimation error for the X parameter e_x would be high. If the E2 neurons contained no information about the X parameter of the stimulus, the X estimation error would be uniformly distributed.

RESULTS

Formation of topographic maps

After training with the Gaussian stimulus set, all layer E1 neurons responded selectively to a small subset of the stimuli. Figure 3*A* shows the response matrix for a typical layer E1 neuron after training with the Gaussian stimulus set. The neuron encodes a continuous subregion of the stimulus space.

To quantify the selectivity, we calculated the mean response for each combination of X and Y stimulus parameter values. To visualize the spatial distribution of the stimulus selectivities, we represented the preferred X and Y parameters of each layer E1 neuron by the hue and the maximal response strength by the brightness of HSV (Hue, Saturation, Value) color space. Figure 4, *A* and *B* shows the topographic maps that were learned with the Gaussian stimuli, using the X dimension as the slow parameter and the Y dimension as the fast changing parameter. Both maps show patches of neighboring neurons with the same or similar selectivities. However, the patches are larger for the X parameter (Fig. 4*A*) than those for the Y parameter (Fig. 4*B*). Moreover, neurons with a preference for the same X parameter are clustered within a single local region of the map. In contrast, patches of neurons with a preference for a certain Y parameter value are distributed across the map.

These properties of the maps are exchanged when the temporal correlations of X and Y parameters are exchanged: Fig. 4, *E* and *F* shows the maps that were learned with the Y dimension as the slowly changing parameter. Here the patches of similar Y preference are larger and localized (Fig. 4*F*), whereas the representation of the X parameter (Fig. 4*E*) shows smaller patches and is more distributed across the map, showing a pattern similar to the pinwheel topography of V1 orientation selectivity (Bonhoeffer and Grinvald 1991). We see that in both cases similar values for the slow parameter (Fig. 4, *A* and *F*) are represented in a localized part of the map, whereas the fast changing parameter has a distributed representation (Fig. 4, *B* and *E*). In many cases the whole range of preferences for the fast changing parameter can be found within a patch of similar preference for the slow parameter.

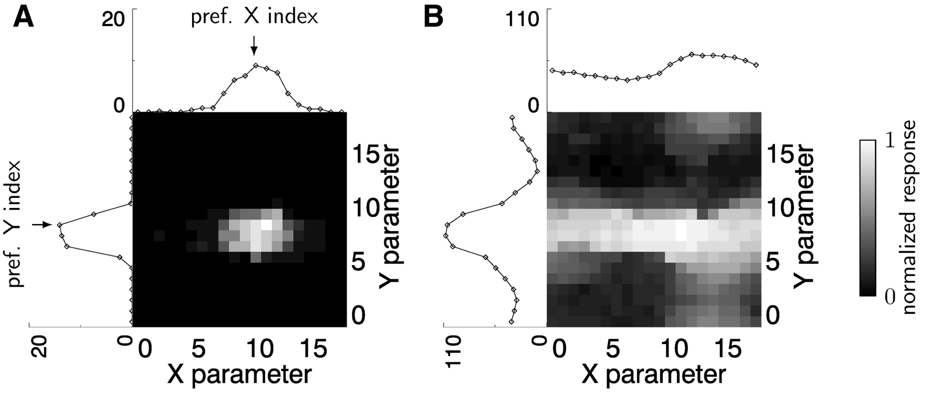


FIG. 3. Learned stimulus selectivities. Stimulus response matrix and X and Y tuning curves for an example layer E1 neuron (*A*) and a layer E2 neuron (*B*). The maxima of the X and Y tuning curves are defined as preferred X and Y indices. The response matrices show the response strength for each of the 400 stimuli. The X and Y tuning curves are a measure for the selectivity to the 2 dimensions of the stimulus set. Here, Y was the slowly changing parameter. *A*: the E1 neuron encodes a subregion of the stimulus space. *B*: response of the E2 neuron showed high selectivity for the Y parameter and low selectivity for the X parameter.

Downloaded from www.physiology.org/journal/jn (084.134.067.044) on May 31, 2019.

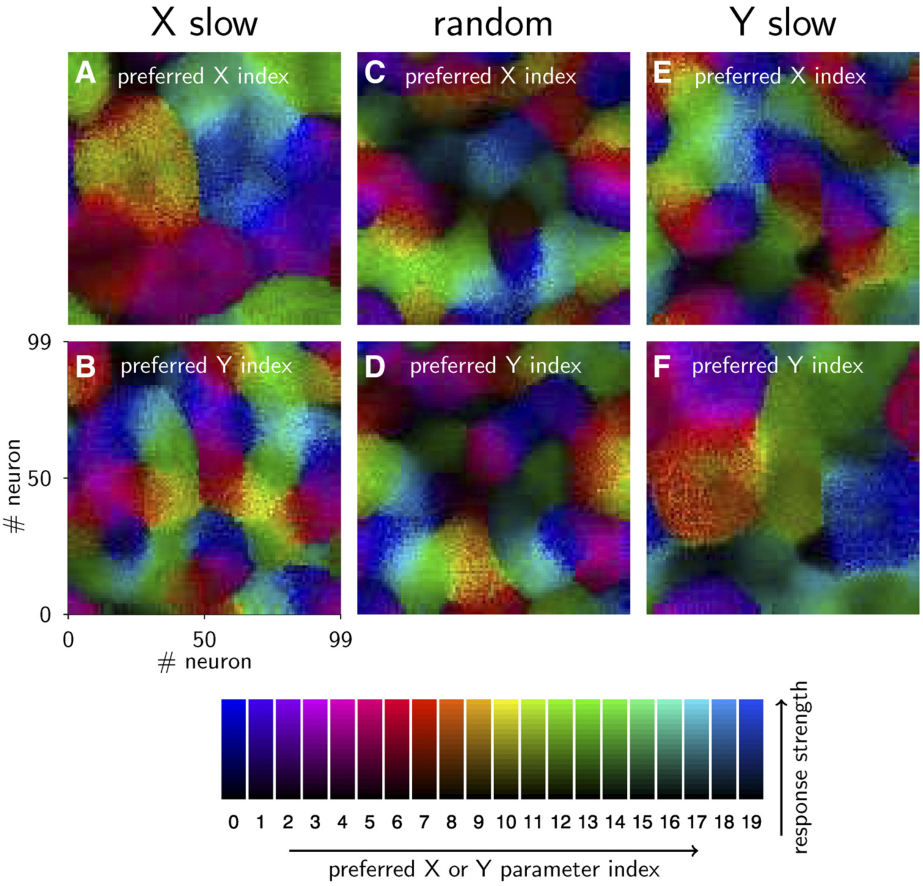


FIG. 4. Learned topographic maps. Preferred X (*A*, *C*, *E*) and Y (*B*, *D*, *F*) stimulus index of layer E1 neurons after learning with Gaussian stimuli. Color indicates preferred parameter values and response strength, as shown by *inset* below panels. The color of each pixel corresponds to the preference value [0–19] of a single layer E1 neuron. Maps are shown for 3 different learning conditions (see *Training and test procedures*). *A* and *B*: X slow. *C* and *D*: random. *E* and *F*: Y slow. The maps for the fast changing parameter (*B* and *E*) have smaller patches and preferences for the same index are distributed across the map. The maps for the slowly changing parameter (*A* and *F*) show larger patches of neurons with similar preference and preferences for similar values are clustered.

In the condition of random presentation (Fig. 4, *C* and *D*), there were no qualitative differences between the maps for the preferred X and Y parameters.

The topographic maps obtained with the prism and COIL stimuli (not shown) looked similar to those of the Gaussian set. To quantitatively compare the patch structure of the different maps, we calculated the Fourier spectra of the topographic maps and used the peak spatial frequency as an estimate of the patch sizes (Table 3). For all simulations with the Gaussian or prism stimuli and X as the slow parameter, the peak spatial frequency for the X parameter was much lower than that for the Y parameter. Conversely, when Y was the slow parameter the peak spatial frequency was lower in the Y map. We can conclude that the topographic maps for the slow parameter show larger patches compared with the maps for the fast changing parameter. This indicates that the temporal correlations are reflected in the learned topography. For the COIL stimuli, in the X slow condition the difference in patch sizes is very small. This is caused by the strong asymmetry in spatial correlations between X and Y dimensions of the stimulus space: strong correlations in the X direction (same object, different viewing angle), low correlations in the Y direction (same viewing angle, different object).

TABLE 3. *Spatial frequencies of topographic maps (Fig. 4), normalized to the dimensions of layer E1 (100 × 100)*

	X Slow		Y Slow	
Stimulus Set	X s.f.	Y s.f.	X s.f.	Y s.f.
Gaussian	0.98	2.16	2.11	0.98
Prism	1.00	2.11	1.62	1.10
COIL	1.76	1.87	1.95	0.98

In all cases the dominant spatial frequency (s.f.) is lower for the slow and higher for the fast and continuously changing parameter. For the COIL stimulus in the X slow condition the differences are very small because of the biased correlation structure of this stimulus set (see text).

To illustrate the topographic order, we determined the regions in the map activated by the same object for different viewing angles (Fig. 5). A patch of high neural activity is continuously shifted as the viewing angle of the object changes, similar to activity in inferotemporal cortex evoked by different views of a face (Wang et al. 1996). Different views of the same object are mapped in the same region and have overlapping representations.

Stability of learned preference maps

To investigate the convergence of the learned representations we performed an analysis of the temporal development of the learned preference maps in a simulation with 20 training epochs of 250 s and with Y as the slow parameter. For each neuron, we calculated the differences between X and Y preference values in each epoch to the respective preference values after the following training epoch. The fraction of neurons with a difference >1 decreased from 62% to 11% for the X preference and from 31% to 4.5% for the Y preference. Both

Downloaded from www.physiology.org/journal/jn (084.134.067.044) on May 31, 2019.

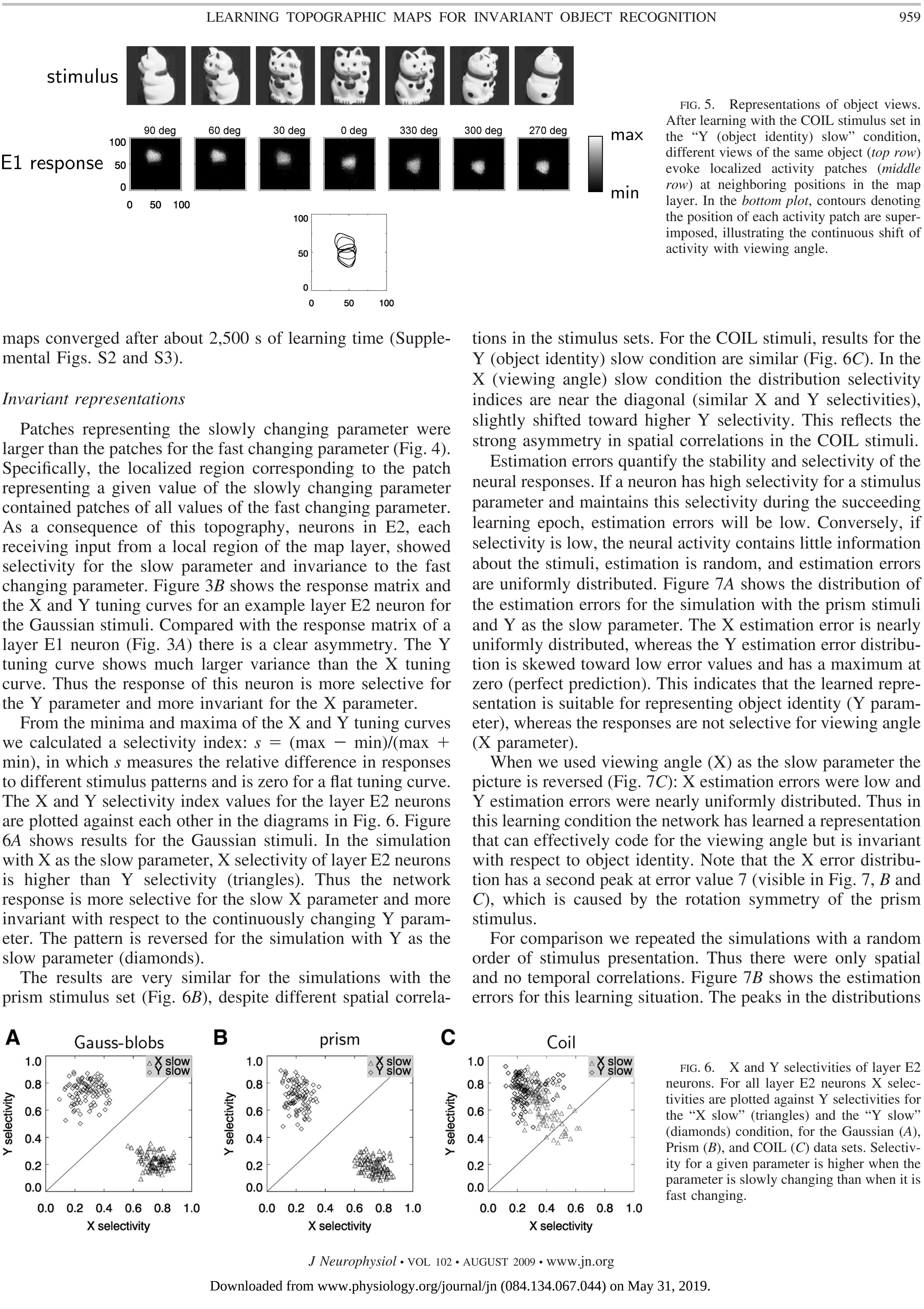


FIG. 5. Representations of object views. After learning with the COIL stimulus set in the "Y (object identity) slow" condition, different views of the same object (*top row*) evoke localized activity patches (*middle row*) at neighboring positions in the map layer. In the *bottom plot*, contours denoting the position of each activity patch are superimposed, illustrating the continuous shift of activity with viewing angle.

maps converged after about 2,500 s of learning time (Supplemental Figs. S2 and S3).

Invariant representations

Patches representing the slowly changing parameter were larger than the patches for the fast changing parameter (Fig. 4). Specifically, the localized region corresponding to the patch representing a given value of the slowly changing parameter contained patches of all values of the fast changing parameter. As a consequence of this topography, neurons in E2, each receiving input from a local region of the map layer, showed selectivity for the slow parameter and invariance to the fast changing parameter. Figure 3*B* shows the response matrix and the X and Y tuning curves for an example layer E2 neuron for the Gaussian stimuli. Compared with the response matrix of a layer E1 neuron (Fig. 3*A*) there is a clear asymmetry. The Y tuning curve shows much larger variance than the X tuning curve. Thus the response of this neuron is more selective for the Y parameter and more invariant for the X parameter.

From the minima and maxima of the X and Y tuning curves we calculated a selectivity index: $s = (\max - \min)/(\max + \min)$, in which s measures the relative difference in responses to different stimulus patterns and is zero for a flat tuning curve. The X and Y selectivity index values for the layer E2 neurons are plotted against each other in the diagrams in Fig. 6. Figure 6*A* shows results for the Gaussian stimuli. In the simulation with X as the slow parameter, X selectivity of layer E2 neurons is higher than Y selectivity (triangles). Thus the network response is more selective for the slow X parameter and more invariant with respect to the continuously changing Y parameter. The pattern is reversed for the simulation with Y as the slow parameter (diamonds).

The results are very similar for the simulations with the prism stimulus set (Fig. 6*B*), despite different spatial correlations in the stimulus sets. For the COIL stimuli, results for the Y (object identity) slow condition are similar (Fig. 6*C*). In the X (viewing angle) slow condition the distribution selectivity indices are near the diagonal (similar X and Y selectivities), slightly shifted toward higher Y selectivity. This reflects the strong asymmetry in spatial correlations in the COIL stimuli.

Estimation errors quantify the stability and selectivity of the neural responses. If a neuron has high selectivity for a stimulus parameter and maintains this selectivity during the succeeding learning epoch, estimation errors will be low. Conversely, if selectivity is low, the neural activity contains little information about the stimuli, estimation is random, and estimation errors are uniformly distributed. Figure 7*A* shows the distribution of the estimation errors for the simulation with the prism stimuli and Y as the slow parameter. The X estimation error is nearly uniformly distributed, whereas the Y estimation error distribution is skewed toward low error values and has a maximum at zero (perfect prediction). This indicates that the learned representation is suitable for representing object identity (Y parameter), whereas the responses are not selective for viewing angle (X parameter).

When we used viewing angle (X) as the slow parameter the picture is reversed (Fig. 7*C*): X estimation errors were low and Y estimation errors were nearly uniformly distributed. Thus in this learning condition the network has learned a representation that can effectively code for the viewing angle but is invariant with respect to object identity. Note that the X error distribution has a second peak at error value 7 (visible in Fig. 7, *B* and *C*), which is caused by the rotation symmetry of the prism stimulus.

For comparison we repeated the simulations with a random order of stimulus presentation. Thus there were only spatial and no temporal correlations. Figure 7*B* shows the estimation errors for this learning situation. The peaks in the distributions

FIG. 6. X and Y selectivities of layer E2 neurons. For all layer E2 neurons X selectivities are plotted against Y selectivities for the "X slow" (triangles) and the "Y slow" (diamonds) condition, for the Gaussian (*A*), Prism (*B*), and COIL (*C*) data sets. Selectivity for a given parameter is higher when the parameter is slowly changing than when it is fast changing.

Downloaded from www.physiology.org/journal/jn (084.134.067.044) on May 31, 2019.

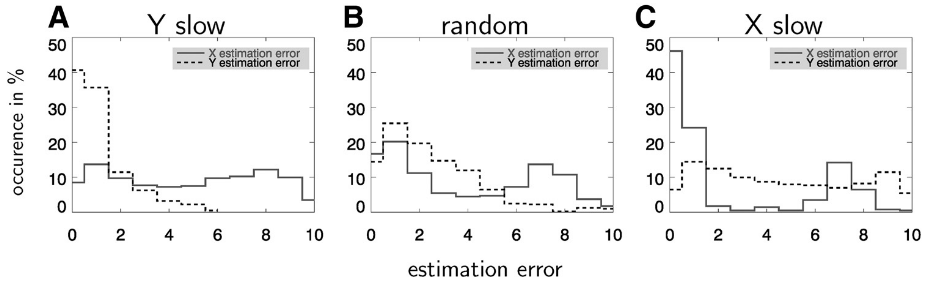


FIG. 7. Estimation errors of layer E2 activity for simulation with the prism stimuli. *A*: Y (object identity) was the slow parameter, X (viewing angle) the fast changing parameter. *B*: random order of stimuli. *C*: X was the slow and Y the fast changing parameter.

for low errors are much smaller and reflect the spatial correlations in the stimuli.

Parameter variations

To test the robustness of the learning mechanism, we systematically varied stimulus timing and the properties of excitatory and inhibitory lateral connections. For these tests we used the Gaussian stimuli with Y as the slow parameter. To evaluate the network performance we defined test trials with an estimation error $e < 2$ as correct predictions and the proportion of correct predictions as the performance. These performance values were plotted against the variations of simulation parameters in Fig. 8. Strong invariance is indicated by a high value in the Y performance and low value in the X performance because Y was the slow parameter. Chance level is $3/20 = 0.15$.

We varied range and strength of lateral excitatory connections ($\sigma_{E1,E1}$, $S_{E1,E1}$), strength of lateral inhibition ($S_{E2,E1}$), and the stimulus timing (t_{stim}). Figure 8, *A–C* shows the dependence of the X and Y performance on the range of the lateral excitatory connections for three different stimulus timing conditions ($t_{stim} \in \{10, 20, 40 \text{ ms}\}$). The network shows high invariance in a range $4 \leq \sigma_{E1,E1} \leq 5$ for all three stimulus timing conditions (Fig. 8, *A–C*). In Fig. 8*D* the stimulus timing was varied in the range $5 \text{ ms} \leq t_{stim} \leq 100 \text{ ms}$, whereas all other parameters were constant. For long stimulus presentation times of $t_{stim} > 70$ ms the performance for the fast and the slow parameters were very similar around 0.5, and thus the responses in layer E2 showed no invariance.

When the strength of the lateral connections between E1 neurons was varied, the network showed high performance in a range $0.07 \leq S_{E1,E1} \leq 0.13$ (Fig. 8*E*). Without the lateral connections ($S_{E1,E1} = 0$), performance dropped to chance level. For weights >0.15 performance decreased as well. Thus although these ranges were fairly broad, lateral extent and strength of the lateral excitation should be within a proper range, corresponding to relative changes by a factor of 2. In contrast, the strength of the lateral inhibitory connections is uncritical (Fig. 8*F*). Network performance is very robust against increased inhibition over a wide range. Likewise, varying the time constants of the learning rule, τ_{pre} and τ_{post}, by a factor of 2 from 10 to 20 ms, did not lead to qualitatively different results (data not shown).

The emergence of topographic maps in our model critically depends on persistent activity in localized groups of neurons, which acts as a memory trace. Figure 9*A* shows how the size of the activity patches representing the stimuli depends on the parameters of the lateral connectivity. Patch size increases with

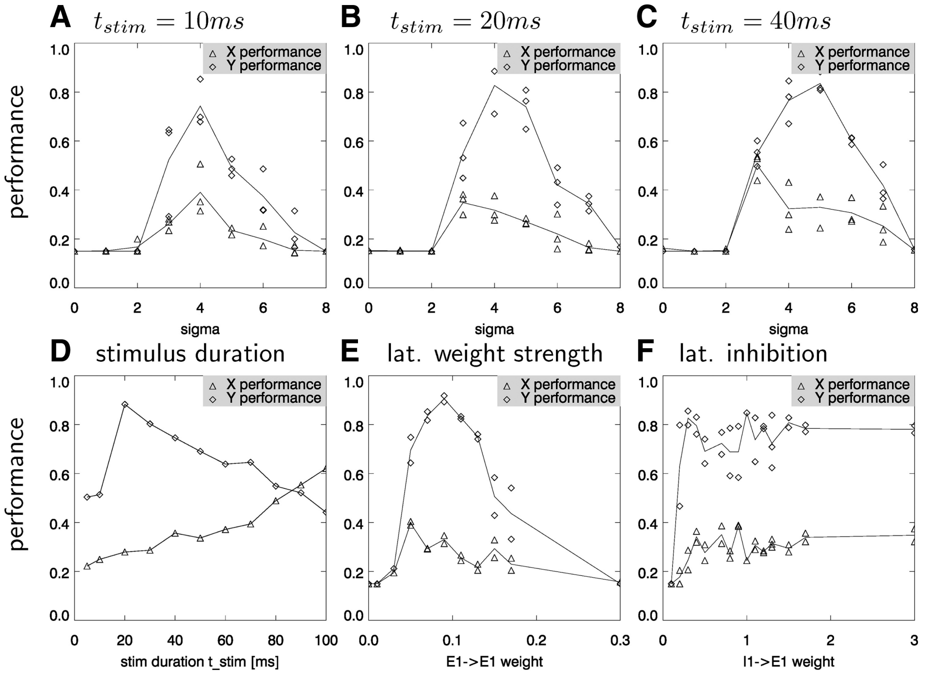


FIG. 8. Effects of model and stimulus parameters on network performance. The network was trained with the Gaussian stimuli with Y as slowly changing parameter. The diagrams show the dependence of the X and Y performance on the range of the lateral connections in E1($\sigma_{E1,E1}$) and stimulus sequence speed t_{stim}. An invariant representation is indicated by high Y performance and low X performance. With a lateral interaction range $\sigma = 4$ the network learned invariant representations for a wide range of stimulus speed values. *A*: fast $t_{stim} = 10$ ms. *B*: $t_{stim} = 20$ ms. *C*: $t_{stim} = 40$ ms. *D*: with increasing stimulus duration t_{stim}, Y performance increases and X performance drops. *E*: strength of the lateral connections in E1($S_{E1,E1}$) was varied. *F*: strength of the lateral inhibitory connections ($S_{I1,E1}$) was varied.

Downloaded from www.physiology.org/journal/jn (084.134.067.044) on May 31, 2019.

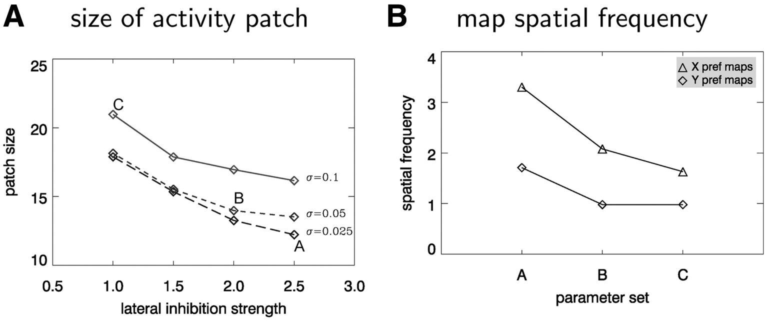


FIG. 9. Patch size depends on range of lateral excitation and strength of inhibition. *A*: size of activity patches plotted against the strength of lateral inhibition $S_{E1,I1}$ for 3 different ranges of lateral excitation $\sigma_{E1,}E_1$. Patch size (in numbers of neurons) is measured as the width at half-height of the activity patch. Patch size increases with larger lateral excitation range $\sigma_{E1,E1}$ and decreases with stronger lateral inhibition. *B*: map spatial frequency for different parameter sets (conditions A–C) in *A*). Larger size of activity patches results in lower spatial frequency of the learned preference map. Y was the slow parameter.

larger lateral excitation range $\sigma_{E1,E1}$ and decreases with stronger lateral inhibition. However, the strength of lateral excitatory connections, determined by the amplitude of the Gaussian kernel $S_{E1,E1}$, did not influence the size of activity patches. Patch size in turn influenced the learned stimulus preference maps. As Fig. 9*B* shows, larger patch size leads to maps with lower spatial frequency.

DISCUSSION

We investigated a mechanism for learning invariant properties of input stimuli. This mechanism implements the idea of extracting slowly varying features from input sequences. It can be applied for learning invariant representations of visual objects. When view-variant retinal projections of an object are presented successively, the spatiotemporal correlations in the input lead to a locally connected, restricted representation in a topographic map. This topographic representation can be used to produce invariant responses in neurons at a successive stage, without further learning, via a simple, unspecific connection scheme. Our approach combines the principles of invariance learning by exploiting temporal correlations and self-organization of topographic maps. Furthermore, it demonstrates that learning of slowly varying features can be achieved in a network of spiking neurons, which is a necessary requirement for a biologically realistic mechanism. Furthermore, our results suggest a functional relevance of cortical topographic maps.

Spatiotemporal input correlations and topographic maps

The architecture of our network is similar to that proposed by von der Malsburg (1973). This architecture is an application of the principle of pattern formation by local self-enhancement and long-range inhibition (Gierer and Meinhardt 1972). The basic building blocks are adaptive, Hebbian forward connections, long-range lateral inhibition, and short-range lateral excitatory connections. Trained with a set of stimuli, such networks transform the spatial correlations between stimuli into spatial proximity of their representations in the emerging map (Choe and Miikkulainen 1997; Kohonen 1982; von der Malsburg 1973).

It is possible to learn view-invariant representations by using spatial correlations only (Stringer et al. 2006), but this requires that spatial correlations between different views of the same object are higher than spatial correlations between views of different objects. This is the case for our simulations with the COIL stimulus set. Even without temporal correlations along the object dimension, the strong spatial correlations along the viewing angle dimension and weak spatial correlations along the object dimension lead to selectivity for object identity. However, in many real-life viewing situations views of different objects (such as faces) can be highly correlated if seen from the same viewing angle, whereas different views of the same object can result in highly different retinal images. With such a stimulus set Wiemer (2003) observed emergence of selectivity for viewing angle. As with our COIL stimulus set, the spatial correlations in the stimulus set dominated the selectivity after learning.

Our prism stimulus set has correlations along both dimensions of the stimulus set (viewing angle and object identity). Under these conditions, spatial correlations alone are not sufficient to learn view-invariant representations that are selective for object identity. Therefore both spatial and temporal correlations must be exploited.

Under natural viewing conditions different views of the same visual object often occur in temporal proximity. We mimicked such viewing conditions by creating stimulus sequences with temporal correlations along only one dimension of the stimulus space. Many different models have been proposed for how these temporal correlations can be used for learning invariant representations of visual objects (Becker 1993; Einhäuser et al. 2005; Földiák 1991; Rolls and Stringer 2006; Stringer and Rolls 2002; Wallis and Rolls 1997; Wiemer 2003; Wiemer et al. 2000; Wiskott and Sejnowski 2002). Our study shows how a biologically plausible network of spiking neurons can make use of temporal correlations to achieve invariant representations.

In contrast to most models of self-organizing maps (e.g., Choe and Miikkulainen 1997; Erwin et al. 1995; Goodhill and Cimponeriu 2000; Goodhill and Willshaw 1990; Kohonen 1982; Swindale 1996; von der Malsburg 1973) in our simulations the network response to a stimulus depends not only on the learned forward connections, but also on the past activity of the map layer. A related principle has been investigated by Wiemer (2003). However, in this study, the relevance of the learned topography for invariant representations was not considered.

Network dynamics and influence of parameters

In previous models for invariance learning from temporal correlations (Einhäuser et al. 2005; Földiák 1991; Rolls and Milward 2000; Wiskott and Sejnowski 2002), the slowness principle was built into the learning rule. In our network, the synaptic learning rule operates only on a fast timescale. It cannot capture temporal correlations on a timescale much longer than 20 ms. Temporal input correlations on a longer

Downloaded from www.physiology.org/journal/jn (084.134.067.044) on May 31, 2019.

timescale are extracted by the network dynamics. Therefore the exact implementation of the learning rule—in particular, the pre- and postsynaptic terms—is uncritical. As Almassy et al. (1998) pointed out, a continuous firing of a local group of neurons has an effect that is similar to Földiák's postsynaptic memory trace. In our network, persistent firing of local groups of E1 neurons is enabled by excitatory lateral interactions in layer E1, which are mediated by fast decaying AMPA currents and slowly decaying NMDA (τ_{decay} = 100 ms, Table 4) currents. These connections provide a local positive feedback, whereas the long-range inhibition reduces the activity in other parts of the layer. This is a neural implementation of the mechanism of biological pattern formation proposed by Gierer and Meinhardt (1972). Note that for this mechanism to work in our case, the time constant of the slow excitatory component (NMDA) must be slower than the time constant of lateral inhibition (GABA). Otherwise, the lateral inhibition would synchronize the whole network and destroy competition between different parts of the map.

The combination of short-range lateral excitatory connections and long-range inhibition enhances activity differences in the E1 layer and results in a competitive network dynamics and local patches of activity can form. Furthermore, in the absence of E0 input, an activated local patch of neurons can keep its activity. This persistent activity is weakened by the depression mechanism in the excitatory lateral synapses. As a result, the patch of activity can move continuously in the E1 layer. Therefore stimuli that occur in temporal sequence—typically different views of the same object—tend to be represented in neighboring regions of the map (Fig. 5).

The specific network dynamics is an essential feature underlying the formation of the topography that captures spatiotemporal correlations. Thus according to our model, one would expect to find persistent activity of local groups of neurons in cortical areas with topographic maps. Furthermore, one would expect that features with similar spatial correlations are represented closer to each other if they are also temporally correlated. This could be tested in experiments investigating the selectivity to object stimuli (e.g., Logothetis et al. 1995) by varying the temporal correlations of the stimuli.

The size of an activity patch in the E1 layer mainly depends on the interaction of positive feedback from the activity center and negative feedback from global inhibition. It increases with longer lateral connections ($\sigma_{E1,E1}$) and decreases by stronger lateral inhibition ($S_{E1,I1}$) (Fig. 9). Despite the dependence of network dynamics on several network parameters, our network is robust against changes in a wide range of parameters (Figs. 8 and 9).

In this study we considered the learning of topographic maps. Other parameters like the connectivity from E1 to E2 were fixed. To achieve invariant responses in layer E2 the convergence from layer E1 to E2 ($\sigma_{E2,E1}$) must be in the range of the patch size in the topographic map for the slow stimulus parameter. Furthermore, we assume that network dynamics and learning rate are appropriate with respect to the typical time constants of changes in the inputs. In a biological network the relevant parameters would have to be adjusted by learning or evolutionary adaptation.

TABLE 4. *Time constants for synaptic currents*

	τ_{raise}, ms	τ_{decay}, ms
AMPA	0.5	2.4
GABA$_A$	1.0	7.0
NMDA	5.5	100.0

Models of invariant representations

An early approach for invariant object recognition is the dynamic routing model (Olshausen et al. 1993). In this model the visual input is transformed into a canonical, object-based reference frame. Although this mechanism can solve the problem of scale and translation invariance, it is insufficient for achieving view invariance because there is no simple geometric transformation between the front view and the back view of an object. Riesenhuber and Poggio (1999) proposed a hierarchical model that relies on the two alternating operations, template matching and pooling units (complex cells), and thereby achieve invariance over the corresponding subset of basic features. They suggest that the proposed connectivity could be learned with the trace rule (Földiák 1991). The VisNet model by Stringer and Rolls (2002) demonstrates how complex-cell connectivity can be learned from temporal correlations in continuous image sequences. Our model extends these approaches and further suggests a possible role of topographic maps for invariant object representations.

Topographic representation and invariant responses

As our results show, a topographic representation can be used to generate invariant responses by simple neural mechanisms. The invariance properties of the output layer (E2) neurons in our model (Fig. 3*B*) are a consequence of the topography in the map layer (E1) because E2 neurons receive input from a localized region in E1 and therefore represent the average activity in this region. After training with sequences of object views, neurons selective for different views of the same object are clustered in a local neighborhood in E1. Neurons in E2 average over such a neighborhood and thus their responses are invariant to viewing angle while maintaining selectivity for object identity. Thus invariance arises from the learned topography through a generic connection scheme without the need for further learning. Without a topography, to achieve invariance in E2 neurons would require specific connections from E1. Learning such specific connections is more costly because a higher number of initial connections must be provided. To achieve an invariant object representation from a population of feature coding cells, those cells must be selected that code for the same object. If these cells were randomly distributed (salt-and-pepper arrangement) in the previous processing layer, a high connectivity would be needed initially to ensure that there is at least one cell in the invariance layer that receives connections from all of them. Furthermore, another learning step would be required to achieve the adequate connectivity for invariant responses. In contrast, in our model invariant responses arise from averaging over a local neighborhood of the topographic map via fixed forward connections that need no further modifications.

The formation of cortical maps has been suggested to be the result of the minimization of wiring length between neurons processing related stimuli (Koulakov and Chklovskii 2001). Our approach is entirely compatible with this view because, in

Downloaded from www.physiology.org/journal/jn (084.134.067.044) on May 31, 2019.

our simulations, the topographic maps emerge as a consequence of the assumption that lateral connections have limited length. In addition, our results demonstrate that the clustering of neurons with similar properties in these maps has the functional benefit that invariance with respect to certain stimulus dimensions can be achieved in a straightforward way.

Conclusions

We propose a mechanism for spatiotemporal correlation-based invariance learning that is compatible with the functional architecture and plasticity mechanisms in the cortex. Our network transforms spatiotemporal correlations of the input sequence into the topography of a self-organizing map. The activity in our network shows similarities to neural activity in inferotemporal cortex (IT), which contains a topographic representation of object features (Tanaka 1996, 2003). The basic mechanisms of our model exist in the ventral pathway of the visual cortex. Therefore it is feasible that the emergence of object feature topography in IT may be based on the principles proposed in our model.

The aim of this work, however, was not to model a specific cortical area. The invariance learning mechanism we described here could be at work for features of any complexity, at any stage in the cortical hierarchy, and in any sensory modality, corresponding to the widely observed occurrence of topographic maps in the cortex.

ACKNOWLEDGMENTS

We thank K. Tanaka for discussions, L. Wiskott and W. Einhäuser for valuable suggestions and critical remarks on an earlier version of the manuscript, and the anonymous reviewers for helpful comments. The Information Technology Department of the University of Marburg provided access to its Modular Architecture for Robust Computation cluster for running simulations.

GRANTS

This work was supported by Deutsche Forschungsgemeinschaft (DFG) Forschergruppe 560 and DFG Graduiertenkolleg 885.

REFERENCES

Almassy N, Edelman GM, Sporns O. Behavioral constraints in the development of neuronal properties: a cortical model embedded in a real-world device. *Cereb Cortex* 8: 346–361, 1998.

Becker S. Learning to categorize objects using temporal coherence. In: *Advances in Neural Information Processing Systems*, edited by Hanson SJ, Cowan JD, Giles CL. San Mateo, CA: Morgan Kaufmann, 1993, vol. 5, p. 361–368.

Bonhoeffer T, Grinvald A. Iso-orientation domains in cat visual cortex are arranged in pinwheel-like patterns. *Nature* 353: 429–431, 1991.

Brunel N, Wang XJ. Effects of neuromodulation in a cortical network model of object working memory dominated by recurrent inhibition. *J Comput Neurosci* 11: 63–85, 2001.

Choe, Y, Miikkulainen R. Self-organization and segmentation with laterally connected spiking neurons. In: *Proceedings of the 15th International Joint Conference on Artificial Intelligence, Nagoya, Japan*. San Francisco, CA: Morgan Kaufmann, 1997, p. 1120–1125.

Choe Y, Miikkulainen R. Self-organization and segmentation in a laterally connected orientation map of spiking neurons. *Neurocomputing* 21: 139–157, 1998.

Crair MC, Malenka RC. A critical period for long-term potentiation at thalamocortical synapses. *Nature* 375: 325–328, 1995.

Deco G, Rolls ET. Neurodynamics of biased competition and cooperation for attention: a model with spiking neurons. *J Neurophysiol* 94: 295–331, 2005.

Einhäuser W, Hipp J, Eggert J, Körner E, König P. Learning viewpoint invariant object representations using a temporal coherence principle. *Biol Cybern* 93: 79–90, 2005.

Einhäuser W, Kayser C, König P, Körding KP. Learning the invariance properties of complex cells from their responses to natural stimuli. *Eur J Neurosci* 15: 475–486, 2002.

Erwin E, Obermayer K, Schulten K. Models of orientation and ocular dominance columns in the visual cortex: a critical comparison. *Neural Comput* 7: 425–468, 1995.

Földiák P. Learning invariance from transformation sequences. *Neural Comput* 3: 194–200, 1991.

Gerstner W, Kempter R, van Hemmen JL, Wagner H. A neuronal learning rule for sub-millisecond temporal coding. *Nature* 383: 76–78, 1996.

Gierer A, Meinhardt H. A theory of biological pattern formation. *Kybernetik* 12: 30–39, 1972.

Goodhill GJ, Cimponeriu A. Analysis of the elastic net model applied to the formation of ocular dominance and orientation columns. *Network* 11: 153–168, 2000.

Goodhill GJ, Willshaw DJ. Application of the elastic net algorithm to the formation of ocular dominance stripes. *Network Comput Neural Syst* 1: 41–59, 1990.

Hubel DH, Wiesel TN. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J Physiol* 160: 106–154, 1962.

Hubel DH, Wiesel TN. Sequence regularity and geometry of orientation columns in the monkey striate cortex. *J Comp Neurol* 158: 267–293, 1974.

Ito M, Tamura H, Fujita I, Tanaka K. Size and position invariance of neuronal responses in monkey inferotemporal cortex. *J Neurophysiol* 19: 218–226, 1995.

Jahr CE, Stevens CF. Voltage dependence of NMDA-activated macroscopic conductances predicted by single-channel kinetics. *J Neurosci* 10: 3178–3182, 1990.

Kohonen T. Self-organized formation of topologically correct feature maps. *Biol Cybern* 43: 59–69, 1982.

Koulakov AA, Chklovskii DB. Orientation preference patterns in mammalian visual cortex: a wire length minimization approach. *Neuron* 29: 519–527, 2001.

Kupper R, Eckhorn R. A neural mechanism for viewing-distance-invariance. In: *Dynamic Perception*, edited by Würtz RP, Lappe M. Berlin: Akademische-Verlag, 2002, p. 277–282.

Logothetis N, Pauls J, Poggio T. Shape representation in the inferior temporal cortex of monkeys. *Curr Biol* 5: 552–563, 1995.

Michler F, Wachtler T, Eckhorn R. Adaptive feedback inhibition improves pattern discrimination learning. In: *Lecture Notes in Computer Science* (*ANNPR*), edited by Schwenker F, Marinai S. New York: Springer, 2006, vol. 4087, p. 21–32.

Miyashita Y. Inferior temporal cortex: where visual perception meets memory. *Annu Rev Neurosci* 16: 245–263, 1993.

Nene S, Nayar S, Murase H. Columbia Object Image Library (COIL-100). Technical Report CUCS-006-96. New York: Columbia Univ. Press, 1996.

Olshausen BA, Anderson CH, Van Essen DC. A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. *J Neurosci* 13: 4700–4719, 1993.

Quian Quiroga R, Reddy L, Kreiman G, Koch C, Fried I. Invariant visual representation by single neurons in the human brain. *Nature* 435: 1102–1107, 2005.

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

Rolls E, Stringer S. Invariant visual object recognition: a model, with lighting invariance. *J Physiol* (*Paris*) 100: 43–62, 2006.

Rolls ET, Milward T. A model of invariant object recognition in the visual system: learning rules, activation functions, lateral inhibition, and information-based performance measures. *Neural Comput* 12: 2547–2572, 2000.

Royer S, Paré D. Conservation of total synaptic weight through balanced synaptic depression and potentiation. *Nature* 422: 518–522, 2003.

Saam M, Eckhorn R. Lateral spike conduction velocity in the visual cortex affects spatial range of synchronization and receptive field size without visual experience: a learning model with spiking neurons. *Biol Cybern* 83: L1–L9, 2000.

Sprekeler H, Michaelis C, Wiskott L. Slowness: an objective for spike-timing-dependent plasticity? *PLoS Comput Biol* 3: e112, 2007.

Stringer SM, Perry G, Rolls ET, Proske JH. Learning invariant object recognition in the visual system with continuous transformations. *Biol Cybern* 94: 128–142, 2006.

Stringer SM, Rolls ET. Invariant object recognition in the visual system with novel views of 3D objects. *Neural Comput* 11: 2585–2596, 2002.

Stryker MP. Temporal associations. *Nature* 354: 108–109, 1991.

Downloaded from www.physiology.org/journal/jn (084.134.067.044) on May 31, 2019.

Swindale NV. The development of topography in the visual cortex: a review of models. *Network* 7: 161–247, 1996.

Tanaka K. Inferotemporal cortex and object vision. *Annu Rev Neurosci* 19: 109–139, 1996.

Tanaka K. Columns for complex visual object features in the inferotemporal cortex: clustering of cells with similar but slightly different stimulus selectivities. *Cereb Cortex* 13: 90–99, 2003.

Tovee MJ, Rolls ET, Azzopardi P. Translation invariance in the responses to faces of single neurons in the temporal visual cortical areas of the alert macaque. *J Neurophysiol* 72: 1049–1060, 1994.

Tsodyks M, Pawelzik K, Markram H. Neural networks with dynamic synapses. *Neural Comput* 10: 821–835, 1998.

Tyberghein J, Zabolotny A, Sunshine E, Hieber T, Galbraith S, Nelson C, Voase M, Wyett P. Crystal Space: Open Source 3D Engine, version 1.0 2007-01-17, Documentation, 2007.

von der Malsburg C. Self-organization of orientation sensitive cells in the striate cortex. *Kybernetik* 14: 85–100, 1973.

Wallis G. Using spatio-temporal correlations to learn invariant object recognition. *Neural Networks* 9: 1513–1519, 1996.

Wallis G, Bülthoff HH. Effects of temporal association on recognition memory. *Proc Natl Acad Sci USA* 98: 4800–4804, 2001.

Wallis G, Rolls ET. Invariant face and object recognition in the visual system. *Prog Neurobiol* 51: 167–194, 1997.

Wang G, Tanaka K, Tanifuji M. Optical imaging of functional organization in the monkey inferotemporal cortex. *Science* 272: 1665–1668, 1996.

Wiemer JC. The time-organized map algorithm: extending the self-organizing map to spatiotemporal signals. *Neural Comput* 15: 1143–1171, 2003.

Wiemer JC, Spengler F, Joublin F, Stagge P, Wacquant S. Learning cortical topography from spatiotemporal stimuli. *Biol Cybern* 82: 173–187, 2000.

Wiskott L, Sejnowski T. Slow feature analysis: unsupervised learning of invariances. *Neural Comput* 14: 715–770, 2002.

Downloaded from www.physiology.org/journal/jn (084.134.067.044) on May 31, 2019.

Supplemental Figures for “Using Spatio-Temporal Correlations to Learn Topographic Maps for Invariant Object Recognition.”

Frank Michler, Reinhard Eckhorn, Thomas Wachtler
Philipps University, Marburg, Germany

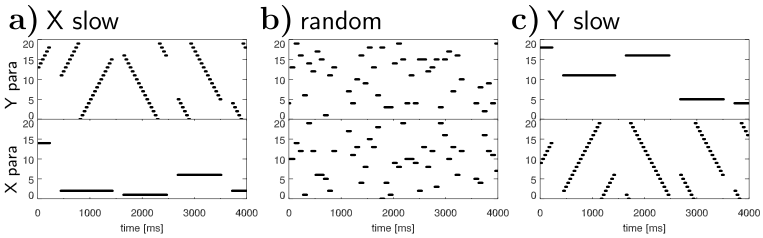


Figure S1: **Stimulus sequences during training.** a) Training condition *“X slow”*. The X parameter was kept constant for time intervals t_{const} (see Methods), while the Y parameter changed continuously. After t_{const}, both parameters switch to randomly chosen values. b) In the “random” training condition there are no temporal correlations between different stimuli c) In the *“Y slow”* training condition the temporal correlations are reversed with respect to the *“X slow”* condition.

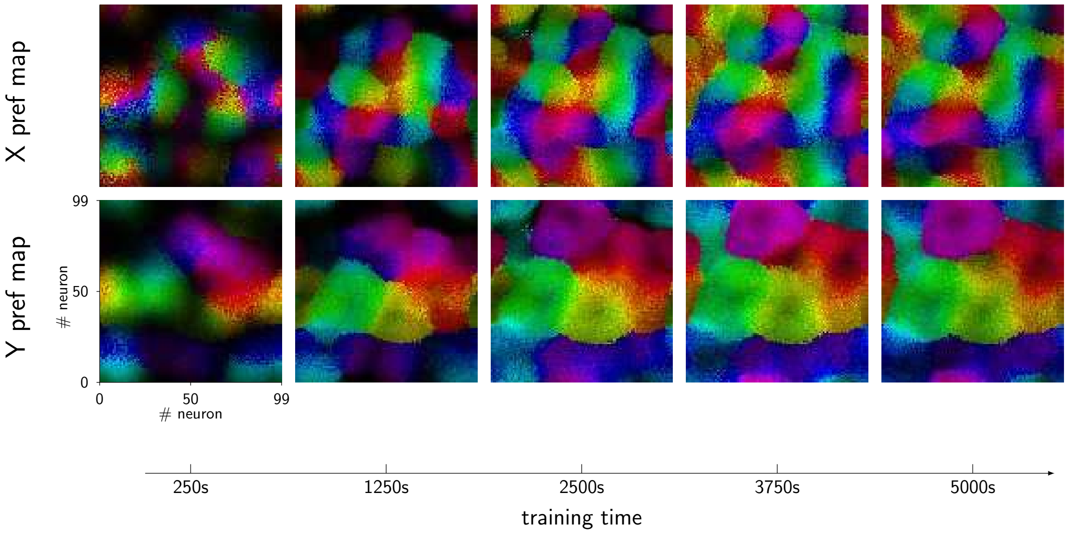


Figure S2: **Time course of X and Y preference maps during learning.** From left to right X (top) and Y (bottom) preference maps are shown after 250s, 1250s, 2500s, 3750s, 5000s training time. Color code as in Figure 4. Stimulus preferences converge after about 2500s training time.

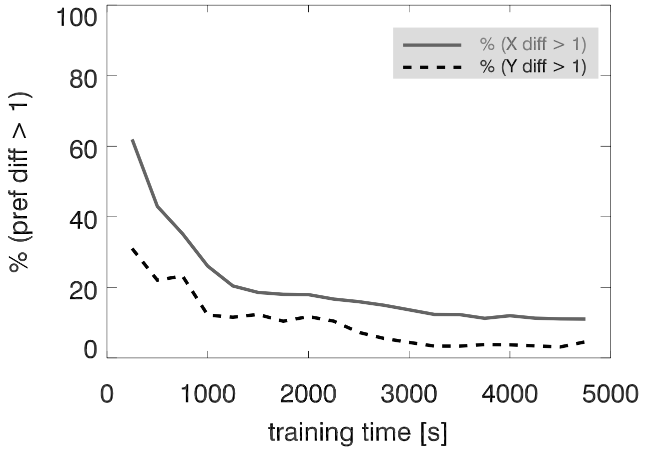


Figure S3: **Convergence of topographic maps.** For each neuron the X and Y preferences after each 250 s training epoch were compared to the preferences after the preceding training epoch. The percentage of neurons with a difference larger than 1 is plotted against learning time for the *"Y slow"* condition. Changes in preferences converge after about 2500 s. The remaining variability is lower for the slow parameter than for the fast changing parameter.

3.2 Adaptive Feedback Inhibition Improves Pattern Discrimination Learning

Summary

The following publication titled “Adaptive Feedback Inhibition Improves Pattern Discrimination Learning” (Michler, Wachtler, and Eckhorn, 2006) addresses the problem of learning to differentiate very similar patterns in a network of spiking neurons. Two well established principles for pattern learning in neural networks are Hebbian plasticity and lateral inhibition. These principles provide the basis for competitive learning, and networks based on them can learn representations suitable to differentiate patterns with a moderate amount of similarity (percentage of overlapping input pixels). However, this solution fails for a set of patterns with a large amount of overlap.

To cope with large overlap, we propose the following mechanism, which implements the idea of predictive coding (Rao and Ballard, 1999) in a network of spiking neurons:

1. Make a reconstruction (prediction) of the input based on the current network activity. This reconstruction represents what the network already "knows" about the current input pattern.
2. Subtract the reconstructed pattern from the actual input.
3. Use the remaining difference to improve the internal representation.

The representation of learned input patterns is encoded in the synaptic weights of feedforward connections from input to output neurons. Subtraction of this known representation can be achieved by inhibitory feedback connections from output to input neurons. Weights for these inhibitory connections are adjusted by an anti Hebbian learning rule.

Our results show that the architecture based on Hebbian learning and lateral inhibition fails to differentiate patterns with an overlap exceeding 75 %. After adding adaptive inhibitory feedback connections, the network learns to differentiate between patterns of up to 88 % overlap.

In conclusion, anti-Hebbian learning of inhibitory feedback connections can improve representations of a feedforward pathway in spiking neural networks.

Declaration of Own Contributions

- All simulations presented in this dissertation were implemented by myself using a C++ based **obj**ect-oriented **sim**ulation library (OBJSIM) for spiking neural networks, which I developed. The source code repository is now published and available along with contributions by Dr. Sebastian Thomas Philipp at https://gin.g-node.org/FrankMichler/ObjSim (Michler and Philipp, 2020).
- I implemented the network architecture for pattern discrimination with adaptive feedback inhibition using OBJSIM.
- I conducted network simulations and parameter scans.
- I developed software for numerical analysis and visualization of simulation results using *Interactive Data Language* (IDL), building upon the vast collection of IDL routines that had been developed in the AG NeuroPhysik Philipps University Marburg.
- I wrote the manuscript in collaboration (discussions, suggestions, editing) with Prof. Dr. Reinhard Eckhorn and Prof. Dr. Thomas Wachtler.
- The article “Adaptive Feedback Inhibition Improves Pattern Discrimination Learning” was peer reviewed by two anonymous reviewers and published as presented here in *Lecture Notes in Computer Science* (Michler, Wachtler, and Eckhorn, 2006).

Adaptive Feedback Inhibition Improves Pattern Discrimination Learning

Frank Michler, Thomas Wachtler, and Reinhard Eckhorn

AppliedPhysics/NeuroPhysics Group, Department of Physics,
Philipps-University Marburg, Renthof 7, D-35032 Marburg, Germany
`frank.michler@physik.uni-marburg.de`
`http://www.physik.uni-marburg.de`

Abstract. Neural network models for unsupervised pattern recognition learning are challenged when the difference between the patterns of the training set is small. The standard neural network architecture for pattern recognition learning consists of adaptive forward connections and lateral inhibition, which provides competition between output neurons. We propose an additional adaptive inhibitory feedback mechanism, to emphasize the difference between training patterns and improve learning. We present an implementation of adaptive feedback inhibition for spiking neural network models, based on spike timing dependent plasticity (STDP). When the inhibitory feedback connections are adjusted using an anti-Hebbian learning rule, feedback inhibition suppresses the redundant activity of input units which code the overlap between similar stimuli. We show, that learning speed and pattern discriminatability can be increased by adding this mechanism to the standard architecture.

1 Introduction

1.1 Standard Architecture

Standard neural networks for unsupervised pattern recognition learning typically consist of adaptive forward connections and lateral inhibition (e.g. Fukushima 1975; Földiák 1990). Usually, the forward connections are modified using Hebbian learning rules: if pre- and postsynaptic activity is highly correlated, excitatory synapses are strengthened while inhibitory synapses are weakened. For excitatory synapses, Hebbian learning increases the correlation between pre- and postsynaptic activity and the connections grow infinitely. Connection strengths can be limited e.g. by using normalization mechanisms.

Lateral inhibitory connections introduce a *winner-take-all (WTA)* dynamics: if an output neuron is strongly activated, other output neurons receive strong inhibition and generate little or no output activity. WTA prevents the output neurons from being active all at the same time. When the lateral inhibitory connections are learned with an *anti-Hebbian* learning rule, as proposed by Földiák (1990), connections are strengthened if correlation between pre- and postsynaptic activity is high. Thus, strongly correlated output neurons will have strong inhibitory connections, which will reduce their correlation. This decorrelation can lead to a sparse representation of the input stimuli (Földiák,

F. Schwenker and S. Marinai (Eds.): ANNPR 2006, LNAI 4087, pp. 21–32, 2006.
© Springer-Verlag Berlin Heidelberg 2006

1990). After self-organization, the neurons in the output layer of such networks should respond selectively to a single stimulus pattern or a subset of the training set, depending on the relation between the size of the stimulus set and the number of output neurons.

1.2 Improving Discrimination Performance with Feedback Inhibition

Consider a two layer network with an input and an output layer, and lateral inhibition between output neurons. What happens when the network is trained with a set of very similar stimuli? Typically the forward connections from the uninformative input neurons coding the overlap between stimuli will become much stronger compared to the connections coding features unique to certain stimuli (Fukushima, 1975; Földiák, 1990). Beyond a certain degree of stimulus similarity the output neurons only respond to the overlap, and thus fail to discriminate between the stimuli. Miyake and Fukushima (1984) proposed a mechanism to improve pattern selectivity fur such situations: they introduced a simple version of modifiable inhibitory feedback connections from the output units to the input units. These connections were paired with modifiable excitatory feedforward connections. When a feedforward connection was strengthened, the corresponding feedback connection was strengthened as well.

In this paper we show that this adaptive feedback inhibition can be generalized and adapted to a biologically more realistic network model with spiking neurons and *spike timing dependent plasticity (STDP)* based learning rule (Bi and Poo, 1998). We systematically varied the overlap between the patterns of the stimulus set and show how learning speed and selectivity increases after introducing modifiable inhibitory feedback connections.

Using spiking neural network models aims towards an understanding of how pattern recognition problems could be solved in the brain. If a mechanism can not be implemented with biologically realistic spiking neurons, then it is unlikely that this mechanism is used in the brain. Furthermore spiking neurons provide for high temporal precision, which is relevant for real-world applications. This is the case e.g. for spatio-temporal pattern recognition or for audio patterns.

2 Model

2.1 Network Architecture

The network is organized in two layers of spiking neurons: the *input layer* U_0 and the *representation layer* U_1 (Fig. 1). There are excitatory forward connections from U_0 to U_1 and lateral inhibitory connections between all U_1 neurons. These connections are adapted due to the correlation between presynaptic and postsynaptic spikes with a Hebbian and anti-Hebbian learning rule, respectively (Section 2.3). So far this is the standard architecture for competitive learning. Additionally, we introduce modifiable inhibitory feedback connections from U_1 to U_0. These inhibitory connections are also adapted using an anti-Hebbian learning rule.

2.2 Model Neurons

As a spiking model neuron we use the two dimensional system of differential equations proposed by Izhikevich (2003):

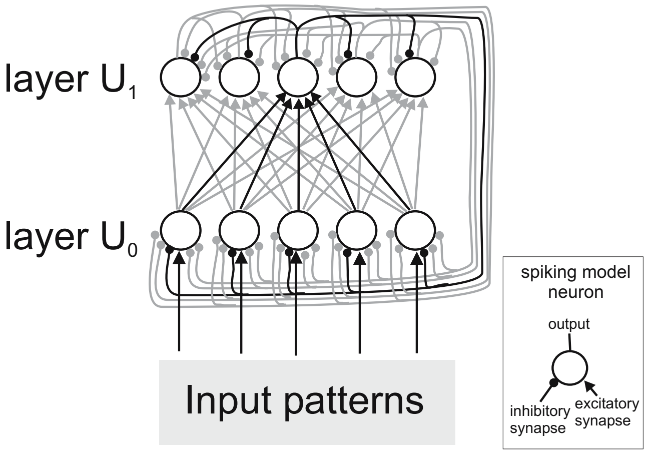


Fig. 1. Model architecture. The neurons of the input layer U_0 are activated when they are part of the current input pattern. U_0 neurons have modifiable excitatory connections to the representation layer U_1. U_1 neurons mutually inhibit each other. Additionally there are modifiable inhibitory feedback connections from U_1 to U_0. To better illustrate the network structure, connections from and to one of the neurons are plotted with black color while the other connections are plotted gray.

$$\frac{dV(t)}{dt} = 0.04V^2(t) + fV(t) + e - U(t) + I(t), \tag{1}$$

$$\frac{dU(t)}{dt} = a(bV(t) - U(t)) \tag{2}$$

with the auxiliary after-spike reseting:

$$\text{if } V(t) \geq 30mV, \text{ then } \begin{cases} V(t) \leftarrow c, \\ U(t) \leftarrow U(t) + d. \end{cases} \tag{3}$$

$V(t)$ and $U(t)$ are dimensionless variables. $V(t)$ represents the membrane potentials in mV. $I(t)$ is the synaptic input current. a, b, c, d, e and f are dimensionless parameters which determine the properties of the model neuron. In the simulations presented here we use a set of parameters which correspond to regular spiking cortical pyramidal neurons (example "L" in Izhikevich, 2004, a=0.02, b=-0.1, c=-55, d=6, e=108, f=4.1). The excitatory synaptic input I_e is modelled as a current injection with additional noise $\sigma(t)$. The inhibitory input I_i is modelled as a conductance based current. The excitatory synaptic input saturates at $I_{e,max}$. The inhibitory conductance saturates at $G_{i,max}$:

$$I = S_e(I_e) - S_i(G_i)(V - E_i), \tag{4}$$

$$S_e(I_e) = I_{e,max}\frac{I_e}{I_e + 1}, \tag{5}$$

$$S_i(G_i) = G_{i,max}\frac{G_i}{G_i + 1}, \tag{6}$$

$$\frac{d}{dt}I_e = -\frac{1}{\tau_e}I_e + \sum_{m=0}^{M-1} w_{e,m}\delta_m(t) + \sigma(t), \tag{7}$$

$$\frac{d}{dt}G_i = -\frac{1}{\tau_i}G_i + \sum_{m=0}^{M-1} w_{i,m}\delta_m(t). \tag{8}$$

The saturation constants were set to $I_{e,max} = 200$ and $G_{i,max} = 4.5$ to restrict excitatory and inhibitory input to a range where the numerical integration of the differential equations still works properly for $dt = 0.25ms$. The excitatory and inhibitory synaptic currents decrease exponentially with time constant τ_e and τ_i respectively, which were arbitrarily set to $5ms$. The biologically realistic range for the decay time constants of excitatory $AMPA$- and inhibitory $GABA_A$-currents is from 5 up to 50 ms. $w_{e,m}$ is the excitatory weight from the presynaptic neuron number m. $\delta_m(t)$ is 1 when a spike arrives at the presynaptic site, otherwise it is 0. E_i is the reverse potential for the inhibitory current which was chosen to be 10 mV lower then the resting potential.

2.3 Learning Rules

The synaptic weight $w_{m,n}$ of the connection from presynaptic U_0 neuron m to postsynaptic U_1 neuron n is adapted according to a Hebbian learning rule:

$$\frac{d}{dt}w_{m,n} = \delta_n(t) R L_{pre,m} L_{post,n}, \tag{9}$$

$$L_{pre,m} = \sum_{t_{sm}} e^{-\frac{t-t_{sm}}{\tau_{pre}}}, \tag{10}$$

$$L_{post,n} = \sum_{t_{sn}} e^{-\frac{t-t_{sn}}{\tau_{post}}}. \tag{11}$$

$\delta_n(t)$ is 1 when a spike occurs in the postsynaptic neuron n. t_{sm} and t_{sn} denote the times of the past pre- and postsynaptic spikes. When a spike occurs, the pre- or postsynaptic *learning potentials* $L_{pre,m}$ or $L_{post,n}$ are increased by 1. They exponentially decrease with time constant $\tau_{pre} = 20ms$ and $\tau_{post} = 10ms$. R is a constant corresponding to the learning rate and was tuned to allow for a weight change between 5 and 20 % after 10 stimulus presentations. For the excitatory connections from layer U_0 to U_1, we use a quadratic normalization rule:

$$w_{m,n}(t) = W \frac{w_{m,n}(t-dt)}{\sqrt{\sum_{m=0}^{M-1} w_{m,n}^2(t-dt)}}, \tag{12}$$

where W is a constant value to adjust the quadratic weight sum. This prevents infinite growing of weights and introduces competition between the input synapses of a postsynaptic neuron. Physiological evidence for the existence of such heterosynaptic interactions were found, e.g., by Royer and Paré (2003). W was set to a value which ensured a medium response activity at the beginning of the learning phase.

For the inhibitory connections we use the following anti-Hebbian learning rule:

$$\frac{d}{dt}w_{m,n} = R\big(\delta_n(t)L_{pre,m} - C\delta_m(t)w_{m,n}L_{post,n}\big), \tag{13}$$

$$L_{pre,m} = e^{-\frac{t-t_{sm}}{\tau_{pre}}}, \tag{14}$$

$$L_{post,n} = e^{-\frac{t-t_{sn}}{\tau_{post}}}. \tag{15}$$

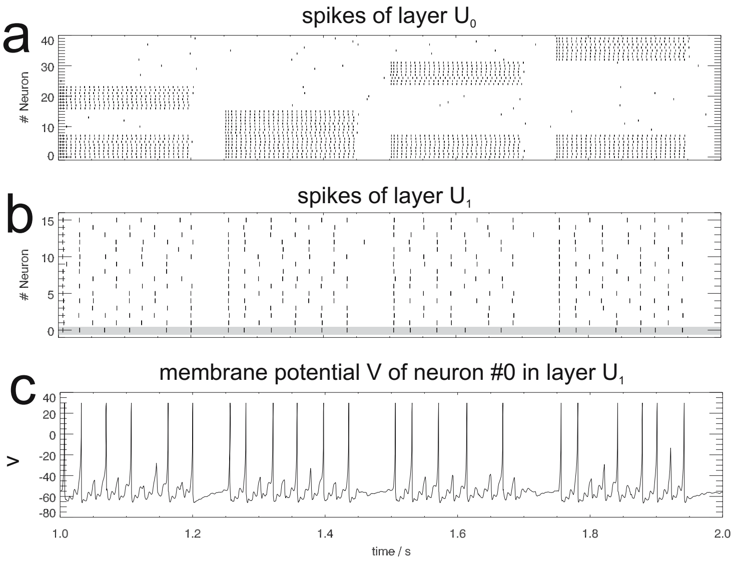


Fig. 2. Network without feedback inhibition, response before learning. a: Spikes of input layer U_0. b: Spikes of representation layer U_1. c: Membrane potential $V(t)$ of neuron #0 of U_1 (gray line in b).

The equations are very similar to the Hebbian learning rule (equation 9) but with an additional depression term. The decay time constants of the learning potentials were set to $\tau_{pre} = 30ms$ and $\tau_{post} = 100ms$. C is a constant to adjust the ratio between potentiation and depression which determines the amount of inhibition. With lower C the inhibitory connections will be stronger. C was set to 0.005 for the feedback inhibition and 0.001 for the lateral inhibition. t_{sm} and t_{sn} denote the time of the last pre- and post- synaptic spike event respectively.

2.4 Stimuli

The input stimuli are binary spatial patterns that lead to additive modulation of excitatory synaptic current I_e (equation 4) of layer U_0 neurons:

$$I_e(t) = \sum_{i \in N} p_n^{k_i} I_0 rect\left(\frac{t - i\tau_1}{\tau_2}\right), \tag{16}$$

$$rect(t) = \begin{cases} 1 : |t| < 0.5 \\ 0 : otherwise\ . \end{cases} \tag{17}$$

$p_{k_i}^n$ is 1 if the neuron n is active for stimulus k_i, and 0 otherwise. I_0 is the input strength. τ_1 is the time difference between stimulus onsets, τ_2 is the duration of a single stimulus presentation (see Fig. 2 for an example). $k_1, k_2, ..., k_i$ is a random sequence of stimulus numbers.

For a systematic variation of the similarity between the input patterns, we constructed sets of stimuli as follows: each stimulus is a binary pattern P_k of N_{U_0} elements where N_{U_0} is the number of neurons in the input layer.

$$P_k = (p_1^k, p_2^k, p_3^k, ..., p_{N_{U_0}}^k), \tag{18}$$

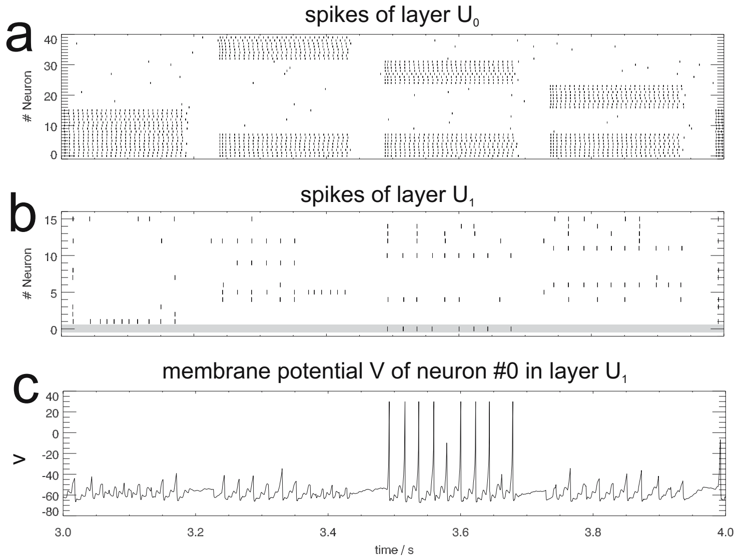


Fig. 3. Network without feedback inhibition, response after learning. a: Spikes of input layer U_0. b: Spikes of representation layer U_1. c: Membrane potential $V(t)$ of neuron #0 of U_1 (gray line in b).

$$p_m^k = \begin{cases} 1 \; , m \leq n_o \\ 1 \; , n_o + n_u(m-1) < m \leq n_o + n_u m \\ 0 \; , otherwise \; . \end{cases} \tag{19}$$

$n_a = f_a N_{U_0}$ is the number and f_a the fraction of active neurons in each pattern. $n_o = f_o n_a$ is the number of neurons which are active in each pattern (overlap) and $n_u = n_a - n_o$ is the number of neurons which are unique for each pattern.

2.5 Performance Measure

In order to quantify the ability of the network to discriminate between the stimuli, we simulated a test phase after every learning phase. In the test phases the network was stimulated with the same input patterns as in the learning phases. We calculated the preferred stimulus κ_n and a selectivity index η_n for every U_1 neuron:

$$\kappa_n = \big\{k : R_{n,k} = max(\{R_{n,1}, ..., R_{1,K}\})\big\}, \tag{20}$$

$$\eta_n = \frac{R_{n,\kappa_n}}{\sum_{k=0}^{K} R_{n,k}} - \frac{1}{K} \; . \tag{21}$$

K is the number of stimuli. κ_n is the number of the stimulus which evokes the maximal response in U_1 neuron n. The selectivity index η_n is 0 if all stimuli evoke the same response $R_{n,k}$, which means that this neuron bears no information about the identity of the stimulus. The maximum selectivity is $\frac{K-1}{K}$ when only one stimulus evokes a response but the others do not. From the following test phase we calculated how the activity of the U_1 neurons predict the identity of the input patterns: for each stimulus onset we derived the response $r_{n,i}$ for every U_1 neuron (number of spikes in a specified interval after stimulus onset), where j is the number of the current stimulus. Combining

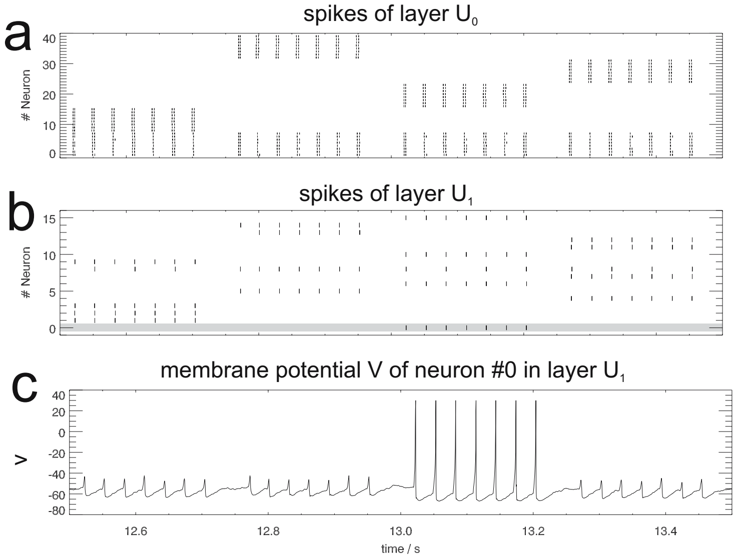


Fig. 4. Network with feedback inhibition, response after learning. a: Spikes of input layer U_0. b: Spikes of representation layer U_1. c: Membrane potential $V(t)$ of neuron #0 of U_1 (gray line in b). The feedback inhibition circuit causes rhythmic spike patterns in both layers.

these responses with the preference and the selectivity of the neurons, we calculated the stimulus ν_j predicted by this network activity:

$$\nu_j = \{k : \xi_k = max(\{\xi_1, ..., \xi_K\})\}, \tag{22}$$

$$\xi_k = \sum_{n \epsilon \{i : \kappa_i = k\}} \eta_n r_{n,k} \, . \tag{23}$$

If $\nu_j = j$ then the prediction is correct, otherwise it is false. The performance ρ is then $\rho = \frac{n_{hit}}{n_{hit}+n_{fail}}$ where n_{hit} is the number of correct predictions and n_{fail} the number of mistakes. The chance level is $\frac{1}{K}$.

3 Results

First we demonstrate the properties of the network without feedback inhibition for a stimulus set with little overlap (50%). The number of stimuli was $K = 4$. The numbers of neurons were: $N_{U_0} = 40$ and $N_{U_1} = 16$. Before learning, the network responds unselectively to the input stimuli (Fig. 2). The network quickly converges to a selective state: for each stimulus there is at least one U_1 neuron that selectively responds to it (Fig. 3).

When we systematically increased the overlap between the elements of the stimulus set the network needed longer to reach a selective state. When the overlap was very high it completely failed to discriminate between the stimuli (Fig. 5).

When we added the modifiable inhibitory feedback connections, the network took less time steps to reach a selective state. Even for high overlap, where it had failed without feedback inhibition, the network learned a selective representations (Fig. 6). Furthermore, the feedback inhibition causes rhythmic spike patterns in both layers and synchronizes the activated neurons (Fig. 4).

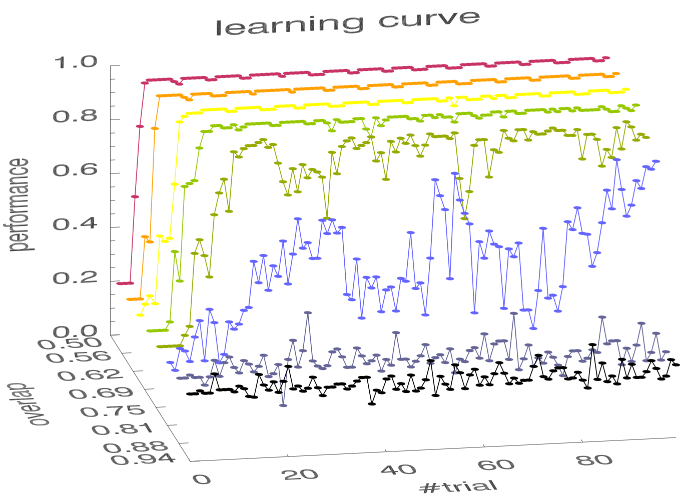


Fig. 5. Learning curves without feedback inhibition. A *trial* consisted of 40 stimulus presentations. For overlap up to 75% the network quickly learned a selective representation. For higher overlap it took longer training time to reach a selective state. For overlap higher than 88% the network stayed in an unselective state. Input strength: $I_0 = 0.008$.

Because the feedback inhibition reduces the spiking activity in U_0, we compensated this effect by increasing excitatory input strength I_0 (see equation 16) when turning on the feedback inhibition. To make sure that the differences in learning speed and learning performance were not caused by these parameter changes, we systematically tested the effect of different input strengths. We calculated a performance index for each I_0 value by averaging the performance values for the second half of learning trials over all overlap levels. Without feedback inhibition the maximum performance of the network (at $I_0 \approx 0.008$) was still lower than the maximum performance of the network with feedback inhibition (Fig. 7).

4 Discussion

Our simulations show that in a network of spiking neurons adaptive feedback inhibition can speed up learning of selective responses and enable discrimination of very similar input stimuli. The mechanism works as follows: While the network is in an unselective state, the correlation between the output units and these input units which code the overlap ($p_1^k ... p_{n_o}^k$ in Eq. 18) is higher than the correlation between the output units and the input units which are unique for different patterns. Therefore, the inhibitory connections to the input neurons representing the overlap will grow stronger and the redundant activity will be reduced. In contrast, the input neurons coding the difference between the stimuli receive less inhibition. Thus, the network can use the discriminative information carried by these neurons to learn a selective representation.

The network parameters were chosen in a biologically realistic range. The input strength I_0 and the feed forward weight sum W were set to obtain reasonable firing rates. The learning parameters that control the inhibitory connections (C, τ_{pre}, τ_{post}) must be guanrantee a substantial amount of inhibition. Overall the mechanism doesn't

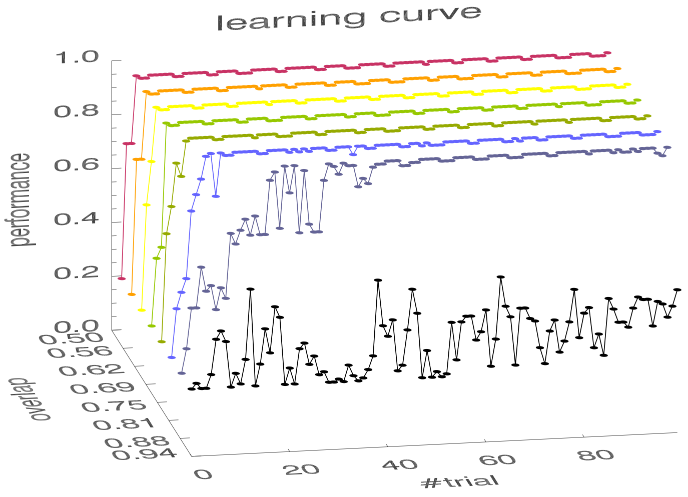


Fig. 6. Learning curves with feedback inhibition ; a *trial* consisted of 40 stimulus presentations. For the low overlap stimulus sets (50% - 81%) the network converged to a selective state faster than without feedback inhibition. Even for very high overlap (94%) the network still learned some selectivity. Input strength: $I_0 = 0.016$.

depend on the precise values of the parameters. Small or medium parameter changes do not qualitatively alter the properties of the network.

4.1 Comparison to Other Models

Miyake and Fukushima (1984) had already proposed a inhibitory feedback mechanism and showed how it could be included in their Cognitron model. They demonstrated the increased selectivity using stimulus pairs with up to 50% spatial overlap. As our simulations show, such an amount of overlap can still be separated using a network without feedback inhibition (Fig. 5).

Spratling (1999) had proposed a *pre-integration lateral inhibition* model. In this model for example an output neuron O_i which has strong excitatory connection from input neuron I_j will have strong inhibitory influence on the excitatory connections from I_j to the other output neurons $O_{k \neq i}$. Spratling and Johnson (2002) showed that *pre-integration lateral inhibition* can enhance unsupervised learning. Spratling (1999) argues against the feedback inhibition model, that an output neuron cannot entirely inhibit the input to all other neurons without entirely inhibiting its own input. van Ooyen and Nienhuis (1993) point out a similar argument: With feedback inhibition the Cognitron model fails to elicit sustained responses for familiar patterns, because the corresponding input activity is deleted. But these drawbacks do not hold in our dynamic model: After strong activation of an output neuron O_i, the feedback inhibition will suppress the input and thus prevent all output neurons from firing including O_i. Inhibition is reduced, and excitatory input can grow again. Thus, for sustained input, the inhibitory feedback generates rhythmic chopping of both input and output layer neurons (Fig. 4). The strongest activated output neurons are able to fire output spikes before inhibition grows, while weakly activated output neurons are kept subthreshold. Furthermore, the

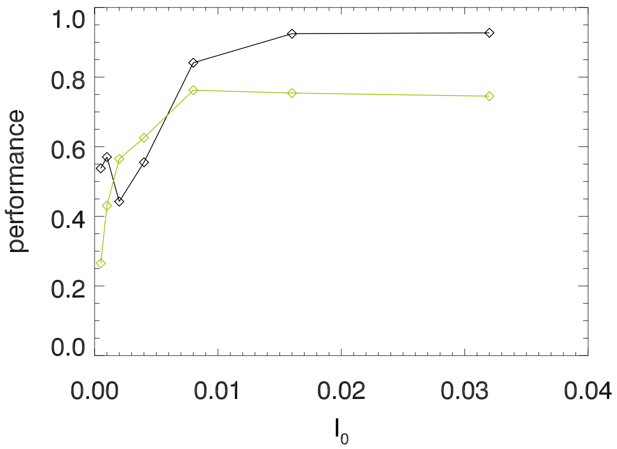


Fig. 7. Performance depends on input strength I_0. The data points show mean performance values, averaged over all overlap values and the second half of the learning trials. Black: Performance with feedback inhibition. Green (gray): Performance without feedback inhibition. Note that with feedback inhibition the network reaches higher performance values (90% compared to 75%).

common feedback inhibition tends to synchronize the activity of these input neurons which are part of the recognized pattern. Such a synchronization has been proposed to support object recognition through dynamic grouping of visual features (see e.g. Eckhorn, 1999; Eckhorn et al., 2004). In the model presented here, synchronization occurs as a consequence of successful pattern recognition.

The adaptive feedback inhibition model is in line with predictive coding models (Rao and Ballard, 1997). These models are based on the working principle of extended Kalman filters, where a prediction signal is subtracted from the input. Thus, in these models the predicted (expected) information is suppressed. This approach is the opposite to the *Adaptive-Resonance-Theorie (ART)*, which is based on enhancement of predicted information (Grossberg, 2001).

4.2 Physiological Equivalent

What could be a physiological basis for the proposed feedback inhibition mechanism? The main input to a cortical area arrives in layer 4 (Callaway, 1998). For example, layer 4 of the primary visual cortex receives input from the thalamic relay neurons of the lateral geniculate nucleus (LGN). Neurons in layer 2/3 have more complex receptive fields. They represent the main output of a cortical module to other cortical areas (Callaway, 1998). Thus, layer U_0 of our model corresponds to cortical layer 4 and layer U_1 to cortical layer 2/3.

Among direct input from thalamic relay neurons, layer 6 neurons receive feedback connections from layer 2/3. In visual area V1 they project back to the LGN but also have collaterals which project to layer 4, where they mainly target inhibitory interneurons (Beierlein et al., 2003). Thus, the anatomy of the neocortex provides the necessary connections for adaptive feedback inhibition: *layer 4 → layer 2/3 → layer 6 → inhibitory*

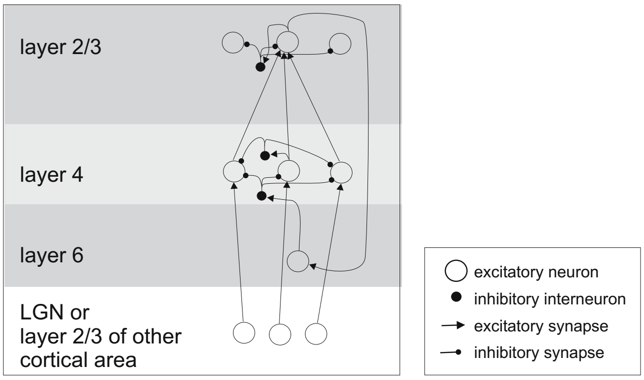


Fig. 8. Possible microcircuit underlying selective feedback inhibition: information enters the cortical module via layer 4, layer 2/3 learns selective representation of input patterns and projects back to layer 6, layer 6 neurons have projections to inhibitory interneurons in layer 4

interneurons of layer 4. This microcircuit could provide the basis for the suppression of uninformative input activity (Fig. 8).

We have shown, that adaptive feedback inhibition can increase learning speed and improve discrimination of highly similar patterns. For simplicity, we used a small set of simple stimulus patterns. The proposed mechanism can also be used for recognition of more complex patterns (e.g. 3d visual objects), if it is incorporated in a hierarchical multi-layer network architecture with feedback inhibition from higher to lower layers.

Acknowledgements

This work was supported by DFG grant EC 53/11.

References

Beierlein, M., Gibson, J. R., Connors, B. W. (2003). Two dynamically distinct inhibitory networks in layer 4 of the neocortex. Journal of Neurophysiology 90, 2987–3000.

Bi, G., Poo, M. (1998). Synaptic modifications in cultured hippocampal neurons: Dependence on spike timing, synaptic strength, and postsynaptic cell type. The Journal of Neuroscience 18 (24), 10464–10472.

Callaway, E. M. (1998). Local circuits in primary visual cortex of the macaque monkey. Annual Review of Neuroscience 21, 47–74.

Eckhorn, R. (1999). Neural mechanisms of scene segmentation: Recordins from the visual cortex suggest basic circuits for linking field models. IEEE Transactions on Neural Networks 10 (3), 464–479.

Eckhorn, R., Bruns, A., Gabriel, A., Al-Shaikhli, B., Saam, M. (2004). Different types of signal coupling in the visual cortex related to neural mechanisms of associative processing and perception. IEEE Transactions on Neural Networks 15 (5), 1039–1052.

Fukushima, K. (1975). Cognitron: A self-organizing multilayered neural network. Biological Cybernetics 20, 121–136.

Földiák, P. (1990). Forming sparse representations by local anti-hebbian learning. Biological Cybernetics 64, 165–170.

Grossberg, S. (2001). Linking the laminar circuits of visual cortex to visual perception: Development, grouping and attention. Neuroscience and Biobeavioral Revies 25, 513–526.

Izhikevich, E. M. (2003). Simple model of spiking neurons. IEEE Transactions on Neural Networks 14 (6), 1569–1572.

Izhikevich, E. M. (2004). Which model to use for cortical spiking neurons? IEEE Transactions on Neural Networks 15 (5), 1063–1070.

Miyake, S., Fukushima, K. (1984). A neural network model for the mechanism of feature-extraction. A self-organizing network with feedback inhibition. Biological Cybernetics 50, 377–384.

Rao, R. P. N., Ballard, D. H. (1997). Dynamic model of visual recognition predicts neural response properties in the visual cortex. Neural Computation 9, 721–763.

Royer, S., Paré, D. (2003). Conservation of total synaptic weight through balanced synaptic depression and potentiation. Nature 422, 518–522.

Spratling, M. W. (1999). Pre-synaptic lateral inhibition provides a better arcitecture for self-organizing neural networks. Network: Computation in Neural Systems 10, 285–301.

Spratling, M. W., Johnson, M. H. (2002). Pre-integration lateral inhibition enhances unsupervised learning. Neural Computation 14 (9), 2157–2179.

van Ooyen, A., Nienhuis, B. (1993). Pattern recognition in the neocognitron is improved by neuronal adaptation. Biological Cybernetics 70, 47–53.

Chapter 4

Discussion

For object recognition it is necessary to discriminate between very similar visual patterns, but also to decide, whether two similar but slightly different patterns represent different views of the same object or different objects. In this dissertation I have put forward four hypotheses addressing these challenges in spiking neural networks.

Sustained Neural Activity can Serve as a Trace Rule

First, I proposed that sustained neural activity can serve as a trace rule for invariance learning. In our model (Michler, Eckhorn, and Wachtler, 2009), sustained firing of neurons in the map layer E1 is enabled by short-range lateral connections with excitatory AMPA- and NMDA-mediated synapses. Lateral inhibition contained this activity to localized activity peaks. Before learning, feedforward synapses (E0 to E1) were initialized with equal weights, and activity dynamics in layer E1 was driven internally. As learning progressed, E1 neurons gained selectivity for presented input patterns. Because the activity peak moved continuously across the map layer, successive input patterns tended to be represented within a local neighborhood. This is similar to the effect of a synaptic trace rule, which binds temporally correlated patterns to the same output neuron.

Topographic Maps can Represent Temporal Correlations

The second hypothesis was that topographic maps can represent temporal correlations. To test this, the network was trained with stimulus sets that were designed with homogeneous spatial correlations along two axis in a 2D feature space. By switching temporal correlations from one axis to the other, effects of temporal correlations on learned topographic maps could be analyzed. Results showed that selectivity patches for the continuously and fast changing feature dimension were smaller than patches for the "slow" dimension. After training the network on the same stimulus set but with switched axis of temporal correlation, the selectivity pattern switched too, proving that neighborhood relations in the learned maps represent not only spatial but also temporal correlations.

Topographic Maps can Enable Invariance

Third, I hypothesized that topographically ordered representations of object views could enable invariant response properties, thereby facilitating invariant object recognition. Indeed, neurons in layer E2, which pooled over local neighborhoods in E1, showed high selectivity to the "slow" stimulus dimension and invariance with respect to the continuously changing dimension.

Adaptive Feedback Inhibition can Improve Learning

The fourth hypothesis was that adaptive feedback inhibition (AFI) can improve discrimination learning for very similar stimuli. A comparison of learned representations for stimulus sets with increasing degree of overlap in chapter 3.2 showed that with AFI the network could discriminate patterns with higher degree of overlap than without AFI. Furthermore, with AFI fewer training trials were needed to learn stable representations.

4.1 Invariant Object Recognition

A comparison of the different approaches for learning the underlying connectivity reveals that, despite the improvements in object recognition performance, we still lack a good understanding of the learning processes that are used in the brain to build invariant object representations. Models of spiking neural networks (SNNs), like the ones presented in this dissertation, can improve our insights into how learning occurs in biological neural networks.

Advances in Computer Vision

While it is easy for humans to invariantly identify objects over a large range of viewing conditions, this task was a major stumbling block for computer vision systems (Pinto, Cox, and DiCarlo, 2008). In the last decade, huge progress was made due to improved computer hardware (especially the use of graphic processing units, GPUs) and the popularization of *Convolutional Neural Networks* (CNNs) (Krizhevsky, Sutskever, and Hinton, 2012; Kriegeskorte, 2015). In recent years, CNN-based models have dominated the annual *ImageNet Large Scale Visual Regocnition Challenge* (Russakovsky et al., 2015), in which research groups compete for the best performance in image recognition tasks on the ImageNet dataset (Deng et al., 2009). With larger and deeper models getting better every year, He et al. (2015) were the first to surpass the performance of a human expert. A year later they further improved and set a new record (He et al., 2016).

In these models, invariance is achieved through alternating template matching and pooling operations (section 1.3, Figure 1.2), similar to the model for simple and complex cells in primary visual cortex (Hubel and Wiesel, 1962). While in our model (Michler, Eckhorn, and Wachtler, 2009) individual map layer neurons perform a similar template matching operation (based on Hebbian learning instead of backpropagation), invariance is achieved in layer E2 neurons by pooling over a local neighborhood of the topographically organized feature map.

Learning Invariance

Whereas the connections in the HMAX model are hard wired, CNNs use backpropagation algorithms, which adjust their weights for filtering and pooling operations to minimize the error of the network output (Werbos, 1990). To calculate this error, the desired output (e.g. the correct label for an input image) must be known. Thus, backpropagation is only possible if large labeled datasets are available to train the network. Although backpropagation has been proven to be an extremely powerful algorithm, it is not considered biologically plausible for a number of reasons (Bengio et al., 2015): First, it is not obvious where the error signal should come from.

Second, there is no biologically plausible mechanism that could propagate the error signal backwards across multiple synapses and neurons.

Hebbian learning rules (section 1.2) provide a biologically plausible mechanism for unsupervised learning of simple cells and higher order feature detectors (like the composite cells in HMAX). For unsupervised learning of connections for the pooling operation, temporal contiguity is utilized by Földiák's trace rule (Földiák, 1991). This has been successfully applied to learn translation invariance (Wallis and Rolls, 1997) and invariance for the viewing angle of 3D objects (Stringer and Rolls, 2002).

Because the trace rule only relies on information that is available locally at the synapse, it is biologically more plausible than backpropagation. While it has been applied successfully in rate-coded neural network models to learn translation invariance (Wallis and Rolls, 1997) and invariance for the viewing angle of 3D objects (Stringer and Rolls, 2002), it is still unclear whether it is suitable for temporal contiguity based learning in spiking neural networks as well.

Gaze-Invariance with Topographic Maps

Philipp (2013) applied the concept of invariance learning with topographic maps to the problem of gaze-invariance. He used a network architecture with a map formation layer similar to Michler, Eckhorn, and Wachtler (2009). In Philipp's model, the map layer received input from two sources: a retinotopic layer and a layer coding the gaze direction. The map layer learned representations that enable the output layer to signal the presence of an object in head-centered coordinates.

4.2 Trace Learning in Spiking Neural Networks

Is there a biological equivalent of Földiák's memory trace (Földiák, 1991) that could enable temporal contiguity based learning in the brain? One possible answer is that the memory trace is directly built into specialized types of synapses, as proposed by Evans and Stringer (2012). A second possibility is that the intrinsic network activity could provide a memory trace as in our model analyzed in section 3.1 (Michler, Eckhorn, and Wachtler, 2009).

Evans and Stringer (2012) implemented trace learning by using a long time constant of 150 ms for excitatory synaptic conductances. This can be interpreted as glutamergic synapses with exclusively NMDA receptors and no AMPA receptors (even though the voltage dependence of NMDA conductances was not modeled; compare equation 7 in Michler, Eckhorn, and Wachtler, 2009). This results in very high firing rates of approximately 200 Hz, which Evans and Stringer describe as being "towards the edge" of the biologically plausible range. When an output neuron is already selective for a stimulus A_1 but not for stimulus A_2, this long time constant will cause the neuron to continue firing after the input switches from A_1 to A_2. Therefore, synaptic weights from input A_2 to this output neuron will be strengthened, and in the future the neuron will also respond to stimulus A_2. Assuming A_1 and A_2 represent different transformations of the same object A, after learning, the output neuron responds to A invariantly with respect to that transformation.

Stringer et al. (2006) have demonstrated that a continuum of spatial correlations between object views can also be exploited for learning of invariant representations. This mechanism is referred to as *continuous transformation* (CT) learning. To separate trace learning from effects of CT learning, Evans and Stringer excluded spatial correlations by using stimuli without any overlap. Therefore, it remains open how their model would cope with considerable spatial correlations between individual stimuli, which is a challenge in realistic object recognition tasks. This question could be answered by training their network with stimulus sets that separate the effects of temporal and spatial correlations (Figure 2, page 26).

For stimulus sets with balanced spatial correlations, changing the temporal order of stimuli during learning significantly changed the learned topographic maps. However, using a stimulus set with strong spatial correlations along one feature dimension only (object identity), spatial correlations dominated the learned maps (COIL stimulus set in Figure 6 C, page 29).

The invariance mechanism in our model also relies on neurons sustaining their activity after a stimulus, which is achieved via excitatory input from lateral connections. A major difference to the model by Evans and Stringer is that different transformations of the same object will not be bound to the same neuron, but to neurons within the local neighborhood. In this way, each E_1 neuron is highly selective for a single stimulus (e.g. a viewing angle of a specific object). Invariance emerges by topographically mapping views of the same object onto a local neighborhood in E_1 and E_2 neurons pooling over these local neighborhoods.

4.3 Sustained Intrinsic Activity

In our model for invariance learning (chapter 3.1), formation of topographic maps in layer E1 relied on persistent activity of local groups of neurons. In the initial stages of learning, this persistent activity slowly moved across layer E1 in a random walk. Temporally correlated input patterns were therefore likely to be represented by nearby neurons. For this mechanism to work properly, the balance between forward input from layer E0 and lateral recurrent input from other E1 neurons is critical.

If lateral connections between E1 neurons are too strong, layer E1 is dominated by its intrinsic persistent activity. Therefore, forward connections from the input layer have no effect, and E1 neurons do not become selective for trained input patterns. On the other hand, if lateral connections are too weak, E1 neurons will not exhibit persistent activity, and temporal correlations are not captured in the learned maps.

Urbanczik and Senn (2014) proposed a synaptic learning rule based on a dendritic prediction error. Instead of using a point neuron model, they simulated a somatic and a dendritic compartment. Their rule adjusts the weight of dendritic synapses in such a way that the dendritic membrane potential predicts the somatic firing rate. They showed that formation of topographic maps is possible when using somatic synapses for lateral connections and dendritic synapses for plastic forward connections from input neurons. When they trained this network with a stimulus set that consists of three clusters of correlated patterns, the network learned topographic maps that reflected these spatial correlations. Because lateral somatic connections only had a weak *nudging* effect on the somatic membrane potential, they did not induce persistent activity.

It would be interesting to test if this learning rule could also be utilized to learn topographic maps that reflect temporal correlations. Instead of persistent activity, longer delays in lateral connections could be used to map presented input patterns to neighboring neurons.

4.4 Empirical Evidence for the Role of Temporal Contiguity

A core feature of the model presented in chapter 3.1 is that temporal contiguity in input sequences is utilized to associate different views of the same object. Psychophysical experiments inspired by this idea found evidence that temporal contiguity also plays a role for face recognition in humans. When views of faces are presented in rapid sequences, response times were faster compared to slow sequences (180 vs. 720 ms per view, Arnold and Sieroff, 2012).

Temporal Smoothness Improves Object Representations

By fully controlling and systematically manipulating the visual environment of newborn chickens, Wood and Wood (2018) evaluated the relationship between temporal smoothness and learning of invariant object representations. Chickens were raised within a "controlled-rearing chamber" where views of virtual 3D objects were presented during the first week of their life. In one condition ("smooth") the viewing angle of the virtual objects changed continuously, whereas in the other ("non-smooth") views were presented in a scrambled order. They found that newborn chickens developed more abstract object representations when exposed to temporally smooth objects. This experimental setup was focused on the aspect of continuous transformation learning (Stringer et al., 2006).

Because in most training conditions used in our studies spatiotemporal smoothness was not excluded, but controlled (by changing the axis of temporal proximity: "X slow" vs "Y slow"), results of Wood and Wood (2018) are not strictly comparable to our model predictions. The training conditions most similar to their experimental design are simulations using the COIL data set as shown in section 3.1 (page 29, Figure 6 C). In the "Y slow" condition with continuously changing viewing angles (corresponding to the "smooth" condition in Wood and Wood, 2018), layer E2 responses were object selective (high "Y" selectivity), whereas in the "X slow" condition, selectivity indices were near the diagonal, indicating a less object specific abstract representation. However, the "X slow" condition does not exactly match their "non-smooth" condition, because in our simulations not one, but many objects were used for training, and views of the same object were not scrambled, but views of other objects were presented between views of the same object. In future simulations, our network could be trained with the same stimulus design as used by Wood and Wood (2018) in order to compare learned representations of the model with their experimental results and to untangle effects of temporal proximity from those of spatiotemporal continuity.

Temporal Proximity vs Spatiotemporal Correlations

Tian and Grill-Spector (2015) conducted a series of psychophysical experiments with the goal to separate contributions of temporal proximity and spatiotemporal continuity to the formation of invariant object representations. In an unsupervised training phase, participants saw views of novel 3D objects either in random order (temporal proximity condition) or in a sequence resembling a continuously rotating object (spatiotemporal continuity condition). Object views spanned a 180° view space, with neighboring views either 7.5° (high similarity condition) or 30° apart (low similarity condition). In a test phase, participants were shown pairs of object views and decided whether or not the images showed the same object. In one series of experiments, test views were identical to the views used in training (known view condition). In another, test views were in between trained views, 3.75° (for high similarity), or 7.5° or 15° (for low similarity) away from the nearest trained view (novel view condition).

When trained with high similarity and tested with known views, there was no advantage in the condition with spatiotemporal continuity compared to temporal proximity. This result is consistent with predictions of continuous transformation learning Stringer et al. (2006). When tested with novel views, similarity between trained views had a significant influence. In the high similarity condition, recognition performance after training with spatiotemporal continuity did not change significantly compared to the performance after training with temporal proximity. However, after training with low similarity, performance was better for the spatiotemporal continuity condition. This suggests that spatiotemporal correlations support learning of representations that enable recognition of interpolated views in between learned views.

The stimulus paradigms used by Tian and Grill-Spector (2015) could be applied to our model to compare their psychophysical results with the emerging properties in our network. If our model was trained with low similarity between neighboring object views, in the spatiotemporal continuity condition I would expect that neighboring views of the same object would be represented nearby within an object patch. However, in the temporal proximity condition with random order of views of the same object, I would expect that on average, neighboring object views are represented further apart in the topographic map. This would cause lower recognition performance for novel test stimuli that are in between learned views, consistent with their experimental findings. I expect this because the novel view would weakly activate representations of neighboring trained views. In the spatiotemporal continuity condition, these weakly activated representations would be nearby within the topographic map and therefore have stronger mutual support through recurrent short-range excitatory lateral connections. In the temporal proximity condition, weakly activated representations would be further apart and therefore have less mutual support. As a consequence, activation of corresponding object invariant neurons in layer E2 would be weaker, and recognition performance should decrease.

Tian and Grill-Spector (2015) hypothesized that ”spatiotemporal continuity might provide broader view tuning compared to temporal proximity.” As described above, a representation based on topographic maps could explain these broader tuning curves.

4.5 Adaptive Feedback Inhibition and Predictive Coding

The theory of *predictive coding* assumes that the brain does not passively respond to sensory inputs, but predicts what should come next based on what it has learned from past regularities (Rao and Ballard, 1999). In line with this theory, Alink et al. (2010) have found reduced responses for predictable stimuli in the primary visual cortex using functional magnetic resonance imaging.

The Adaptive Feedback Inhibition (AFI) model presented in chapter 3.2 demonstrates how, in a network of spiking neurons, inhibitory feedback connections that are adjusted by spike-timing dependent plasticity (STDP) can speed up learning and improve internal representations of trained stimuli. This is a biologically plausible implementation of one aspect of predictive coding: subtraction of a prediction from the actual input. Feedback signals from a higher area of feature detectors to a lower level area can be interpreted as a prediction or reconstruction of detected patterns. When inhibitory connections suppress input activity that corresponds to already learned patterns, non-matching parts of the input become more salient and can be learned faster. This model shows how a biologically plausible implementation of predictive coding is possible, and thereby available for learning in the brain.

Hierarchical Models for Predictive Coding

As a proof-of-principle, only a small set of very simple generic patterns was used in our study. For recognition of more complex and realistic patterns, the AFI mechanism could be incorporated in a hierarchical multi-layer network architecture. An example of predictive coding in a hierarchical network architecture are autoencoders, which generate a prediction of the input from an internal representation and use the difference to guide learning (Hinton and Salakhutdinov, 2006). Despite the fact that this type of network is also trained using the backpropagation algorithm, it does not need huge labeled data sets to learn useful object representations as is the case for CNNs. Instead, the difference between pixel pattern in the output and input layer is used as an error signal. When trained with natural images, such networks can learn sparse representations similar to those found in the visual cortex (Vincent et al., 2010).

Whereas autoencoders use the difference between input and output as an error signal to adjust weights in all layers of the hierarchy, a biologically more plausible approach is to calculate a prediction error in each layer of the hierarchy. Spratling (2017) showed how predictive coding in a two-stage hierarchical network can be applied to problems like recognition of hand-written letters.

The examples reviewed so far apply the predictive coding principle to static inputs. In natural viewing situations, input patterns change over time, and consecutive inputs are correlated. In a network using the AFI mechanism, the feedback activity generated by past input patterns would coincide with current inputs. Feedback connections adjusted with an STDP-based learning rule could therefore learn to predict temporal changes in input patterns. Lotter, Kreiman, and Cox (2016) demonstrated how a network, only optimized to predict future frames of video sequences ("PredNet"), develops internal representations suitable for invariant object recognition.

Predictive Coding in the Auditory System

Several experimental studies have shown effects that are consistent with the assumption that predictive coding plays a crucial role in sensory processing. In electroencephalography studies of the auditory system, a phenomenon known as mismatch negativity (MMN) was observed (Näätänen and Alho, 1995). The MMN is an enhanced response that can be measured when an unexpected "deviant" auditory event is occasionally inserted into a repetitive series of "standard" auditory stimuli. A model of the auditory cortex, based on predictive coding, accounts for critical features of the MMN (Wacongne, Changeux, and Dehaene, 2012). Similar to the microcircuit proposed by us (Figure 8 in Michler, Wachtler, and Eckhorn, 2006), Wacongne, Changeux, and Dehaene (2012) used activity of layer 2/3 pyramidal neurons as prediction signals. They interpreted the interaction of excitatory feedforward input and inhibitory feedback in layer 4 as a calculation of a prediction error. To enable predictions based on past stimuli, layer 2/3 neurons are connected to a short-term memory module that keeps a trace of past activity.

Minimizing Free Energy

Friston (2010) has put predictive coding in the context of minimizing "free energy". In this conceptual framework, free energy is related to the amount of surprise about sensory input. By adjusting the internal model about the causes of sensory input in a way that sensory input can be "explained away" (predicted) by the internal representation, surprise and therefore free energy is minimized. As an example, let us assume that two similar input patterns are represented by activity of the same output neuron. The overlapping part of both patterns is explained away by the internal representation, whereas the unique part of the actual stimulus is a surprise. Once the two stimuli are represented by two different output neurons, the amount of uncertainty, and thereby free energy, is reduced, because also the unique part of the stimulus patterns is explained away by the internal representation. Adaptive feedback inhibition enhances the surprising part of sensory inputs relative to the predicted part, and higher level internal representations can be adjusted via competition and Hebbian learning.

Activity of Prediction and Error Neurons

In a recent review of empirical evidence for predictive coding, Heilbron and Chait (2018) argue that, according to predictive coding models, activity differences between neurons in superficial and deep cortical layers should be expected. In predictive coding models, forward connections carry the error signal and feedback connections the prediction signal. Whereas forward connections originate from superficial pyramidal neurons (layer 2/3), feedback originates from deep layers (pyramidal neurons in layer 5/6). Therefore, prediction and error computations should have *distinct laminar profiles*. However, of the few studies addressing this issue, one found no activity difference between superficial and deep layers (Szymanski, Garcia-Lazaro, and Schnupp, 2009), and another found that attenuation was much stronger in deep layers (Rummell, Klee, and Sigurdsson, 2016).

Contrary to the assumption by Heilbron and Chait, an implementation of predictive coding with spiking neurons would not necessarily predict stronger attenuation in error neurons compared to prediction neurons. In the adaptive feedback inhibition model (section 3.2), a correct prediction of sensory input by higher level neurons reduces activity in both layers. When U_1 neurons (Figure 1 in Michler, Wachtler, and Eckhorn, 2006) are activated (prediction), U_0 neurons representing sensory input and prediction error are inhibited, thereby cutting off the input for U_1. As a consequence, U_1 activity is reduced as well, reducing inhibition to U_0 neurons so they can start firing again. Thus, feedback inhibition generates oscillations, synchronizes activity in U_0, and reduces the total number of action potentials in both layers.

4.6 Combining AFI and Topographic Map Learning

Stimulus sets for the invariance-learning simulations (chapter 3.1) were designed with an overlap of similar stimuli below 80 % to enable successful discrimination in the map layer E1. This was done to study the temporal correlation based formation of topographic maps in isolation without introducing unnecessary complexity. Further, the input layer dimensions of 20×20 (for Gaussian and prism stimuli) and $24 \times 26 \times 8$ (for COIL stimuli) were small enough to allow full connectivity from input layer E0 to the map formation layer E1. To enable the network to learn invariant representations for more realistic stimulus sets with larger images and higher levels of similarity, several enhancements of the model would be necessary.

First, the adaptive feedback inhibition (AFI) mechanism described in chapter 3.2 could be used: adaptive inhibitory connections can be added from layer E1 to E0 neurons. Second, similar to HMAX (Riesenhuber and Poggio, 1999) and VisNet (Wallis and Rolls, 1997), the architecture of the model could be repeatedly applied within a hierarchy. Connectivity between input layer and E1 neurons would be spatially limited to form localized receptive fields. The pattern of lateral inhibition would have to be adjusted in order to limit competition, so neurons with non-overlapping receptive fields would not inhibit each other. Instead of a single activity peak, the input would then be represented by multiple peaks that are active simultaneously, leading to a parts-based representation, as was also proposed by Hosoda et al. (2009).

Output of the pooling layer E2 could be used as input for the next map layer. Within such a hierarchy, map layer neurons are similar to simple cells (S1 and S2 layers in Figure 1.2), whereas neurons pooling over local neighborhoods of map layers correspond to complex cells (C1 and C2 in Figure 1.2).

Parker and Serre (2015) have extended the HMAX model to enable learning of transformation sequences. They used a "temporal pooling" mechanism to arrange local features of consecutive object views into the same pool of simple cells that constitute the input for the MAX pooling operation of complex cells. The model was trained with movie sequences of rotating objects. After training, they compared sensitivity of the network to non-accidental properties (NAPs) and metric properties (MPs). NAPs correspond to properties that are invariant to viewpoint, e.g. whether an edge is straight or curved. In contrast, MPs change continuously with in-depth rotation, like the length of an edge or the angle between two edges. They found that complex cells showed higher selectivity for NAPs than for MPs, consistent with behavioral and electrophysiological data (Biederman, 2007). I expect a similar selectivity difference for NAPs vs MPs in a hierarchical version of the model proposed in Michler, Eckhorn, and Wachtler (2009), because the "temporal pooling" in this extended HMAX model is similar to the pooling over a local neighborhood of map neurons.

4.7 Why Study Spiking Neural Networks?

The fact that the biological brains operate with spiking neurons is an obvious reason to continue research on spiking neural networks. However, rate-based models (like CNNs) are getting better every year, even surpassing human performance in some tasks like image classification (He et al., 2015) or playing the game of Go (Silver et al., 2017). This begs the question whether there are still other reasons to study spiking neural networks besides the quest to understand the human brain. The fundamental difference between spiking and rate-based neural networks is the mechanism by which information is transmitted between neurons. In rate based models, the output of a neuron must be transmitted to all its efferent neurons in every simulated time interval, whereas in spiking neural networks, a transmission only needs to be processed when the neuron spikes. Because only a small proportion of all neurons are spiking at the same time, while the rest is silent, much less information needs to be exchanged between neurons. This can translate into huge efficiency gains as studies show that implemented neural networks in neuromorphic hardware (Khan et al., 2008; Brüderle et al., 2011; Davies et al., 2018). Research of the capabilities and possible processing mechanisms in spiking neural network is necessary to make use of these new hardware platforms.

Neuromorphic hardware can also be a valuable tool for computational neuroscience research, as it enables simulating of much larger networks than what is possible on classical CPUs. The enhancements I proposed in section 4.6 would increase the number of neurons by many magnitudes (at least by a factor of 10), compared to the model described in Michler, Eckhorn, and Wachtler (2009). Because the number of synapses increases in a superlinear way, simulation times could become so large that experiments with the model would become unfeasible. However, in neuromorphic hardware all neurons and synapses work in parallel (like in the brain), so models can be scaled up and still run fast enough to work with.

4.8 Conclusion

The topographic order of representations in self-organizing maps can be influenced by temporal correlations. Simulations with spiking neural networks have demonstrated how the temporal order of views of visual objects can be encoded in the spatial neighborhood relations within a cortical area. Such topographic maps can emerge from unsupervised learning with Hebbian learning rules that operate on a fast time scale, because sustained firing of local groups of neurons can provide a memory trace, obviating the need for a synaptic trace rule. These results suggest a mechanism that could be responsible for the formation of topographic object representations in the inferotemporal cortex and offer an explanation for their functional role.

Further, plastic inhibitory connections from a higher to a lower level within a neural processing hierarchy can speed up the emergence of accurate representations via unsupervised learning, in line with theories of predictive coding.

The mechanisms described in this dissertation, which are based on temporal learning, topographic representations, and adaptive feedback inhibition, are most likely not exclusive to the visual domain. If so, they can be adopted to other sensory representations as well.

Bibliography

Alink, A. et al. (2010). Stimulus Predictability Reduces Responses in Primary Visual Cortex. In: *The Journal of Neuroscience* 30.8, 2960–2966. DOI: 10.1523/JNEUROSCI.3730-10.2010.

Arnold, G. and Sieroff, E. (2012). Timing constraints of temporal view association in face recognition. In: *Vision research* 54, 61–67. DOI: 10.1016/j.visres.2011.12.001.

Bengio, Y. et al. (2015). Towards Biologically Plausible Deep Learning. In: *Computing Research Repository (CoRR)* abs/1502.04156.

Bi, G. and Poo, M. (1998). Synaptic Modifications in Cultured Hippocampal Neurons: Dependence on Spike Timing, Synaptic Strength, and Postsynaptic Cell Type. In: *The Journal of Neuroscience* 18.24, 10464–10472. DOI: 10.1523/JNEUROSCI.18-24-10464.1998.

Biederman, I. (2007). Recent psychophysical and neural research in shape recognition. In: *Object Recognition, Attention, and Action*. Ed. by N. Osaka, I. Rentschler, and I. Biederman. Tokyo: Springer. ISBN: 978-4-431-73019-4. DOI: 10.1007/978-4-431-73019-4.

Blakemore, C. and Cooper, G. F. (1970). Development of the Brain Depends on the Visual Environment. In: *Nature* 228, 477–478. DOI: 10.1038/228477a0.

Bliss, T. V. and Lømo, T. (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. In: *The Journal of Physiology* 232.2, 331–356. DOI: 10.1113/jphysiol.1973.sp010273.

Bosking, W. H. et al. (1997). Orientation Selectivity and the Arrangement of Horizontal Connections in Tree Shrew Striate Cortex. In: *Journal of Neuroscience* 17.6, 2112–2127. DOI: 10.1523/JNEUROSCI.17-06-02112.1997.

Bower, J. and Beeman, D. (2003). The book of GENESIS: Exploring Realistic Neural Models with the GEneral NEural SImulation System. Springer, New York. ISBN: 978-0-387-94938-3. DOI: 10.1007/978-1-4612-1634-6.

Brüderle, D. et al. (2011). A comprehensive workflow for general-purpose neural modeling with highly configurable neuromorphic hardware systems. In: *Biological Cybernetics* 104.4, 263–296. DOI: 10.1007/s00422-011-0435-9.

Cajal, S. R. y (1894). The Croonian lecture. – La fine structure des centres nerveux. In: *Proceedings of the Royal Society of London* 55, 444–468. DOI: 10.1098/rspl.1894.0063.

Chen, S. et al. (2019). Brain-Inspired Cognitive Model With Attention for Self-Driving Cars. In: *IEEE Transactions on Cognitive and Developmental Systems* 11.1, 13–25. DOI: 10.1109/TCDS.2017.2717451.

Choe, Y. and Miikkulainen, R. (1998). Self-organization and segmentation in a laterally connected orientation map of spiking neurons. In: *Neurocomputing* 21.1, 139–157. DOI: 10.1016/S0925-2312(98)00040-X.

Dan, Y. and Poo, M.-M. (2004). Spike Timing-Dependent Plasticity of Neural Circuits. In: *Neuron* 44.1, 23–30. DOI: 10.1016/j.neuron.2004.09.007.

Dan, Y. and Poo, M.-M. (2006). Spike Timing-Dependent Plasticity: From Synapse to Perception. In: *Physiological Reviews* 86.3, 1033–1048. DOI: `10.1152/physrev.00030.2005`.

Davies, M. et al. (2018). Loihi: A Neuromorphic Manycore Processor with On-Chip Learning. In: *IEEE Micro* 38.1, 82–99. DOI: `10.1109/MM.2018.112130359`.

Delac, K., Grgic, M., and Grgic, S. (2006). Independent comparative study of PCA, ICA, and LDA on the FERET data set. In: *International Journal of Imaging Systems and Technology* 15.5, 252–260. DOI: `10.1002/ima.20059`.

Deng, J. et al. (2009). ImageNet: A Large-Scale Hierarchical Image Database. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–255. DOI: `10.1109/CVPR.2009.5206848`.

Dorosheva, E., Yakovlev, I., and Reznikova, Z. (2011). An Innate Template for Enemy Recognition in Red Wood Ants. In: *Entomological Review* 91.2, 274 –280. DOI: `10.1134/S0013873811020151`.

Evans, B. and Stringer, S. (2012). Transform-invariant visual representations in self-organizing spiking neural networks. In: *Frontiers in Computational Neuroscience* 6.46, 1–19. DOI: `10.3389/fncom.2012.00046`.

Földiák, P. (1991). Learning Invariance from Transformation Sequences. In: *Neural Computation* 3.2, 194–200. DOI: `10.1162/neco.1991.3.2.194`.

Földiák, P. (1992). Models of sensory coding. Tech. rep. CUED/F-INFENG/TR–91. Department of Engineering, University of Cambridge.

Friston, K. (2010). The free-energy principle: a unified brain theory? In: *Nature Reviews Neuroscience* 11, 127. DOI: `10.1038/nrn2787`.

Fukushima, K. (1975). Cognitron: A Self-organizing Multilayered Neural Network. In: *Biological Cybernetics* 20, 121–136. DOI: `10.1007/BF00342633`.

Fukushima, K. (1980). Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position. In: *Biological Cybernetics* 36, 193–202. DOI: `10.1007/bf00344251`.

Gibson, J. (1966). The senses considered as perceptual systems. Oxford, England: Houghton Mifflin.

Gollisch, T. and Meister, M. (2008). Rapid Neural Coding in the Retina with Relative Spike Latencies. In: *Science* 319.5866, 1108–1111. DOI: `10.1126/science.1149639`.

Grossberg, S. (1969). On learning, information, lateral inhibition, and transmitters. In: *Mathematical Biosciences* 4.3, 255–310. DOI: `10.1016/0025-5564(69)90015-7`.

Grossberg, S. (1973). Contour Enhancement, Short Term Memory, and Constancies in Reverberating Neural Networks. In: *Studies in Applied Mathematics* 52.3, 213–257. DOI: `10.1002/sapm1973523213`.

He, K. et al. (2015). Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. In: *2015 IEEE International Conference on Computer Vision (ICCV)*, 1026–1034. DOI: `10.1109/ICCV.2015.123`.

He, K. et al. (2016). Deep Residual Learning for Image Recognition. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778. DOI: `10.1109/CVPR.2016.90`.

Hebb, D. O. (1949). The Organization of Behavior: A neuropsychological theory. New York: John Wiley. ISBN: 978-0805843002.

Heilbron, M. and Chait, M. (2018). Great Expectations: Is there Evidence for Predictive Coding in Auditory Cortex? In: *Neuroscience* 389. Sensory Sequence Processing in the Brain, 54–73. DOI: `10.1016/j.neuroscience.2017.07.061`.

Hinton, G. E. and Salakhutdinov, R. R. (2006). Reducing the Dimensionality of Data with Neural Networks. In: *Science* 313.5786, 504–507. DOI: `10.1126/science.1127647`.

Hodgkin, A. L. and Huxley, A. F. (1952). A Quantitative Description of Membrane Current and its Application to Conduction and Excitation in Nerve. In: *The Journal of Physiology* 117, 500–544. DOI: `10.1113/jphysiol.1952.sp004764`.

Homberg, U. et al. (2011). Central neural coding of sky polarization in insects. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 366.1565, 680–687. DOI: `10.1098/rstb.2010.0199`.

Hosoda, K. et al. (2009). A Model for Learning Topographically Organized Parts-Based Representations of Objects in Visual Cortex: Topographic Nonnegative Matrix Factorization. In: *Neural Computation* 21.9, 2605–2633. DOI: `10.1162/neco.2009.03-08-722`.

Hubel, D. H. and Wiesel, T. N. (1962). Receptive Fields, Binocular Interaction and Functional Architecture in the Cat's Visual Cortex. In: *The Journal of Physiology* 160.1, 106–154. DOI: `10.1113/jphysiol.1962.sp006837`.

Hubel, D. H. and Wiesel, T. N. (1970). The period of susceptibility to the physiological effects of unilateral eye closure in kittens. In: *The Journal of Physiology* 206.2, 419–436. DOI: `10.1113/jphysiol.1970.sp009022`.

Intrator, N. and Edelman, S. (1997). Competitive learning in biological and artificial neural computation. In: *Trends in Cognitive Sciences* 1.7, 268–272. DOI: `10.1016/S1364-6613(97)01066-8`.

Izhikevich, E. M. (2003). Simple Model of Spiking Neurons. In: *IEEE Transactions on Neural Networks* 14.6, 1569–1572. DOI: `10.1109/TNN.2003.820440`.

Izhikevich, E. M. (2004). Which Model to Use for Cortical Spiking Neurons? In: *IEEE Transactions on Neural Networks* 15.5, 1063–1070. DOI: `10.1109/TNN.2004.832719`.

Jameel, M. and Kumar, S. (2018). Handwritten Urdu Characters Recognition Using Multilayer Perceptron. In: *International Journal of Applied Engineering Research* 13.11, 8981–8984.

Kaas, J. H. (1997). Topographic Maps are Fundamental to Sensory Processing. In: *Brain Research Bulletin* 44.2, 107 –112. DOI: `10.1016/S0361-9230(97)00094-4`.

Khan, M. M. et al. (2008). SpiNNaker: Mapping neural networks onto a massively-parallel chip multiprocessor. In: *2008 IEEE International Joint Conference on Neural Networks*, 2849–2856. DOI: `10.1109/IJCNN.2008.4634199`.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. In: *Biological Cybernetics* 43.1, 59–69. DOI: `10.1007/BF00337288`.

Kriegeskorte, N. (2015). Deep Neural Networks: A New Framework for Modeling Biological Vision and Brain Information Processing. In: *Annual Review of Vision Science* 1.1, 417–446. DOI: `10.1146/annurev-vision-082114-035447`.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. In: *Advances in Neural Information Processing Systems 25*. Ed. by F. Pereira et al. Curran Associates, Inc., 1097–1105. DOI: `10.1145/3065386`.

Land, M. F. (1969). Structure of the Retinae of the Principal Eyes of Jumping Spiders (Salticidae: Dendryphantinae) in Relation to Visual Optics. In: *Journal of Experimental Biology* 51.2, 443–470.

Lapicque, L. É. (1907). Recherches quantitatives sur l'excitation électrique des nerfs traitée comme une polarisation. In: *J. Physiol. Pathol. Gen.* 9, 620–635. DOI: `10.1007/s00422-007-0189-6`.

Leaver, A. M. and Rauschecker, J. P. (2016). Functional Topography of Human Auditory Cortex. In: *The Journal of Neuroscience* 36.4, 1416–1428. DOI: `10.1523/JNEUROSCI.0226-15.2016`.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. In: *Nature* 521, 436–444. DOI: `10.1038/nature14539`.

LeCun, Y. et al. (1998). Gradient-based learning applied to document recognition. In: *Proceedings of the IEEE* 86.11, 2278–2324. DOI: `10.1109/5.726791`.

Lettvin, J. Y. et al. (1959). What the Frog's Eye Tells the Frog's Brain. In: *Proceedings of the Institute of Radio Engineers* 47.11, 1940–1951. DOI: `10.1109/JRPROC.1959.287207`.

Lotter, W., Kreiman, G., and Cox, D. (2016). Deep Predictive Coding Networks for Video Prediction and Unsupervised Learning. In: *ArXiv* abs/1605.08104.

Markram, H et al. (1997). Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. In: *Science* 275.5297, 213–215. DOI: `10.1126/science.275.5297.213`.

Michler, F., Eckhorn, R., and Wachtler, T. (2009). Using Spatiotemporal Correlations to Learn Topographic Maps for Invariant Object Recognition. In: *Journal of Neurophysiology* 102.2, 953–964. DOI: `10.1152/jn.90651.2008`.

Michler, F. and Philipp, S. T. (2020). ObjSim. DOI: `10.12751/g-node.00fbef`.

Michler, F., Wachtler, T., and Eckhorn, R. (2006). Adaptive Feedback Inhibition Improves Pattern Discrimination Learning. In: *Artificial Neural Networks in Pattern Recognition*. Ed. by F. Schwenker and S. Marinai. Vol. 4087. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 21–32. ISBN: 3-540-37951-7. DOI: `10.1007/11829898_3`.

Nagaveni, G. and Sreenivasulu Reddy, T. (2014). Detection of an Object by using Principal Component Analysis. In: *International Journal of Engineering Research & Technology* 3.1.

Nath, A. and Schwartz, G. W. (2017). Electrical synapses convey orientation selectivity in the mouse retina. In: *Nature Communications* 8.2025, 1–15. DOI: `10.1038/s41467-017-01980-9`.

Näätänen, R. and Alho, K. (1995). Mismatch negativity: a unique measure of sensory processing in audition. In: *International Journal of Neuroscience* 80, 317–337. DOI: `10.3109/00207459508986107`.

Parker, S. M. and Serre, T. (2015). Unsupervised invariance learning of transformation sequences in a model of object recognition yields selectivity for non-accidental properties. In: *Frontiers in computational neuroscience* 9, 115–115. DOI: `10.3389/fncom.2015.00115`.

Philipp, S. T. (2013). Information Integration and Neural Plasticity in Sensory Processing Investigated at the Levels of Single Neurons, Networks, and Perception. PhD thesis. LMU Munich.

Pinto, N., Cox, D. D., and DiCarlo, J. J. (2008). Why is Real-World Visual Object Recognition Hard? In: *PLOS Computational Biology* 4.1, 1–6. DOI: `10.1371/journal.pcbi.0040027`.

Rall, W. (1959). Branching dendritic trees and motoneuron membrane resistivity. In: *Experimental Neurology* 1.5, 491–527. DOI: `10.1016/0014-4886(59)90046-9`.

Rao, R. P. N. and Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. In: *Nature* 2.1, 79–87. DOI: `10.1038/4580`.

Riesenhuber, M. and Poggio, T. (1999). Hierarchical models of object recognition in cortex. In: *Nature Neuroscience* 2.11, 1019–1025. DOI: `10.1038/14819`.

Rolls, E. T. and Tovee, M. J. (1994). Processing speed in the cerebral cortex and the neurophysiology of visual masking. In: *Proceedings of the Royal Society of London. Series B: Biological Sciences* 257.1348, 9–15. DOI: `10.1098/rspb.1994.0087`.

Rolls, E. T. et al. (1992). Neurophysiological mechanisms underlying face processing within and beyond the temporal cortical visual areas. In: *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 335.1273, 11–21. DOI: `10.1098/rstb.1992.0002`.

Rosenblatt, F. (1958). The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain. In: *Psychological Review* 65.6, 386–408. DOI: `10.1037/h0042519`.

Ruff, H. A., Kohler, C. J., and Haupt, D. L. (1976). Infant recognition of two- and three-dimensional stimuli. In: *Developmental Psychology* 12.5, 455–459. DOI: `10.1037/0012-1649.12.5.455`.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. In: *Nature* 323, 533–536. DOI: `10.1038/323533a0`.

Rummell, B. P., Klee, J. L., and Sigurdsson, T. (2016). Attenuation of Responses to Self-Generated Sounds in Auditory Cortical Neurons. In: *Journal of Neuroscience* 36.47, 12010–12026. DOI: `10.1523/JNEUROSCI.1564-16.2016`.

Russakovsky, O. et al. (2015). ImageNet Large Scale Visual Recognition Challenge. In: *International Journal of Computer Vision* 115.3, 211–252. DOI: `10.1007/s11263-015-0816-y`.

Ryu, J., Yang, M.-H., and Lim, J. (2018). DFT-based Transformation Invariant Pooling Layer for Visual Classification. In: *Computer Vision – ECCV 2018*. Ed. by V. Ferrari et al. Cham: Springer International Publishing, 89–104. ISBN: 978-3-030-01264-9. DOI: `10.1007/978-3-030-01264-9_6`.

Saenz, M. and Langers, D. R. (2014). Tonotopic mapping of human auditory cortex. In: *Hearing Research* 307, 42 –52. DOI: `10.1016/j.heares.2013.07.016`.

San Roque, L. et al. (2015). Vision verbs dominate in conversation across cultures, but the ranking of non-visual verbs varies. In: *Cognitive Linguistics* 26.1, 31–60. DOI: `10.1515/cog-2014-0089`.

Shouval, H. Z., Bear, M. F., and Cooper, L. N. (2002). A unified model of NMDA receptor-dependent bidirectional synaptic plasticity. In: *Proceedings of the National Academy of Sciences* 99.16, 10831–10836. DOI: `10.1073/pnas.152343099`.

Silver, D. et al. (2017). Mastering the game of Go without human knowledge. In: *Nature* 550, 354. DOI: `10.1038/nature2427010.1038/nature24270`.

Simard, P. et al. (1991). Tangent Prop: A Formalism for Specifying Selected Invariances in an Adaptive Network. In: *Proceedings of the 4th International Conference on Neural Information Processing Systems*. NIPS'91. Denver, Colorado: Morgan Kaufmann Publishers Inc., 895–903. ISBN: 1-55860-222-4.

Spratling, M. W. (2017). A Hierarchical Predictive Coding Model of Object Recognition in Natural Images. In: *Cognitive Computation* 9.2, 151–167. DOI: `10.1007/s12559-016-9445-1`.

Stringer, S. M. et al. (2006). Learning invariant object recognition in the visual system with continuous transformations. In: *Biological Cybernetics* 94.2, 128–142. DOI: `10.1007/s00422-005-0030-z`.

Stringer, S. M. and Rolls, E. T. (2002). Invariant Object Recognition in the Visual System with Novel Views of 3D Objects. In: *Neural Computation* 11.14, 2585–2596. DOI: `10.1162/089976602760407982`.

Szymanski, F. D., Garcia-Lazaro, J. A., and Schnupp, J. W. H. (2009). Current Source Density Profiles of Stimulus-Specific Adaptation in Rat Auditory Cortex. In: *Journal of Neurophysiology* 102.3, 1483–1490. DOI: `10.1152/jn.00240.2009`.

Tanaka, K. (1996). Inferotemporal cortex and object vision. In: *Annual Review of Neuroscience* 19, 109–139. DOI: `10.1146/annurev.ne.19.030196.000545`.

Tanaka, K. (2003). Columns for Complex Visual Object Features in the Inferotemporal Cortex: Clustering of Cells with Similar bat Slightly Different Stimulus Selectivities. In: *Cerebral Cortex* 13, 90–99. DOI: `10.1093/cercor/13.1.90`.

Thorpe, S., Fize, D., and Marlot, C. (1996). Speed of processing in the human visual system. In: *Nature* 381.6582, 520–522. DOI: `10.1038/381520a0`.

Tian, M. and Grill-Spector, K. (2015). Spatiotemporal information during unsupervised learning enhances viewpoint invariant object recognition. In: *Journal of Vision* 15.6/7, 1–13. DOI: `10.1167/15.6.7`.

Tsodyks, M, Pawelzik, K, and Markram, H (1998). Neural networks with dynamic synapses. In: *Neural Computation* 10.4, 821–835. DOI: `10.1162/089976698300017502`.

Turrigiano, G. G. and Nelson, S. B. (2004). Homeostatic plasticity in the developing nervous system. In: *Nature Reviews Neuroscience* 5.2, 97–107. DOI: `10.1038/nrn1327`.

Tyberghein, J. et al. (2007). Crystal Space: Open Source 3D Engine Documentation. http://www.crystalspace3d.org/docs/online/manual/index.html.

Urbanczik, R. and Senn, W. (2014). Learning by the Dendritic Prediction of Somatic Spiking. In: *Neuron* 81/3, 521–528. DOI: `10.1016/j.neuron.2013.11.030`.

Vincent, P. et al. (2010). Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion. In: *Journal of Machine Learning Research* 11, 3371–3408.

von der Malsburg, C. (1973). Self-organization of orientation sensitive cells in the striate cortex. In: *Kybernetik* 14.2, 85–100. DOI: `10.1007/BF00288907`.

Wacongne, C., Changeux, J.-P., and Dehaene, S. (2012). A Neuronal Model of Predictive Coding Accounting for the Mismatch Negativity. In: *Journal of Neuroscience* 32.11, 3665–3678. DOI: `10.1523/JNEUROSCI.5003-11.2012`.

Waldrop, M. M. (2015). Autonomous vehicles: No drivers required. In: *Nature* 518.7537, 20–23. DOI: `10.1038/518020a`.

Wallis, G. (1996). Using Spatio-temporal Correlations to Learn Invariant Object Recognition. In: *Neural Networks* 9.9, 1513–1519. DOI: `10.1016/S0893-6080(96)00041-X`.

Wallis, G. and Bülthoff, H. H. (2001). Effects of temporal association on recognition memory. In: *Proceedings of the National Academy of Sciences of the USA* 98.8, 4800–4804. DOI: `10.1073/pnas.071028598`.

Wallis, G. and Rolls, E. T. (1997). Invariant face and object recognition in the visual system. In: *Progress in Neurobiology* 51.2, 167–194. DOI: `10.1016/S0301-0082(96)00054-8`.

Wang, G., Tanaka, K., and Tanifuji, M. (1996). Optical imaging of functional organization in the monkey inferotemporal cortex. In: *Science* 272.5268, 1665–1668. DOI: `10.1126/science.272.5268.1665`.

Werbos, P. J. (1990). Backpropagation through time: what it does and how to do it. In: *Proceedings of the IEEE* 78.10, 1550–1560. DOI: `10.1109/5.58337`.

Werbos, P. (1975). Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences. PhD thesis. Harvard University, Cambridge, MA.

Westheimer, G. (2001). The Fourier Theory of Vision. In: *Perception* 30.5, 531–541. DOI: `10.1068/p3193`.

Wood, J. N. and Wood, S. M. W. (2018). The Development of Invariant Object Recognition Requires Visual Experience With Temporally Smooth Objects. In: *Cognitive Science* 42.4, 1391–1406. DOI: `10.1111/cogs.12595`.

Zenke, F., Hennequin, G., and Gerstner, W. (2013). Synaptic Plasticity in Neural Networks Needs Homeostasis with a Fast Rate Detector. In: *PLOS Computational Biology* 9.11, 1–14. DOI: `10.1371/journal.pcbi.1003330`.

Zhang, R. (2019). Making Convolutional Networks Shift-Invariant Again. In: *CoRR* abs/1904.11486.

Zoellick, J. C. et al. (2019). Assessing acceptance of electric automated vehicles after exposure in a realistic traffic environment. In: *PloS one* 14.5, 1–23. DOI: `10.1371/journal.pone.0215969`.

Danksagung

An dieser Stelle möchte ich noch einmal zurückschauen und all jenen danken, die mich dabei unterstützt haben, diese Dissertation fertigzustellen.

Mein ganz besonderer Dank gilt meinen Betreuern Prof. Dr. Thomas Wachtler und Prof. Dr. Uwe Homberg, die es mir trotz des langen Zeitraums seit Beginn der Arbeit ermöglicht haben, diese abzuschließen. Prof. Dr. Thomas Wachtler danke ich vor allem für die geduldige Betreuung meines Promotionsvorhabens und die produktive Zusammenarbeit bei den Publikationen. Prof. Dr. Reinhard Eckhorn hat mich bis zu seiner Emeritierung betreut und mir in der AG NeuroPhysik die Möglichkeit geboten, in dem spannenden Gebiet der Neurowissenschaften zu forschen. Dafür bin ich ihm zutiefst dankbar.

Auch meinen ehemaligen Kollegen der AG NeuroPhysik sowie allen Mitgliedern des Graduiertenkollegs NeuroAct möchte ich für viele anregende wissenschaftliche Diskussionen und inspirierende Zusammenarbeit danken, ganz besonders Markus Wittenberg, Dr. Timm Zwickel, Dr. Basim Al-Shaikhli und Dr. Sebastian Philipp. Gerne denke ich zurück an spannende politische Diskussionen mit Timm. Basims Schwärmerei für Python war es, die mich dazu angeregt hat, mir diese Sprache auch anzueignen und sehr schnell lieben zu lernen. Sehr dankbar bin ich Basim und Sebastian für ihre spontane Bereitschaft zum Korrekturlesen und die wertvollen Rückmeldungen.

Sarah Schwöbel danke ich für den konstruktiven wissenschaftlichen Austausch während der Zeit ihrer Masterarbeit in München. Bei Dr. Andreas Wolfsteller, Advaita Dick, Christian Schauss und Dr. Teodora Ivanova möchte ich mich für wertvolle Anmerkungen zu dieser Arbeit bedanken. Ganz herzlich bedanke ich mich bei Sylvia Jankowiak für unermüdliches Korrekturlesen, motivierende Gespräche sowie viele hilfreiche Anregungen und Tipps.

Auch Freunde und Familie haben einen großen Anteil daran, dass ich an diesem Projekt festgehalten und es schließlich zu Ende gebracht habe. Meinen Tischtennisfreunden vom FauEffEll danke ich für viele schöne und schweißtreibende Trainingsstunden und Wettkämpfe mit dem kleinen Plastikball. Besonders die Freundschaft mit Alex weiß ich zu schätzen, die in einer Ära begann, als die Bälle noch kleiner und aus Zelluloid waren. Meinen Geschwistern Diana und Andrea bin ich unendlich dankbar dafür, zu wissen, dass wir auch in schwierigen Zeiten immer füreinander da sind. Teodora danke ich für ihre Geduld, Unterstützung und den leckeren Lachs. Unsere bezaubernden Begegnungen auf der Tanzfläche haben mein Leben immens bereichert.

Wissenschaftlicher Werdegang

Diese Seite enthält personenbezogene Daten, die nicht in der elektronisch publizierten Version der Arbeit enthalten sind.