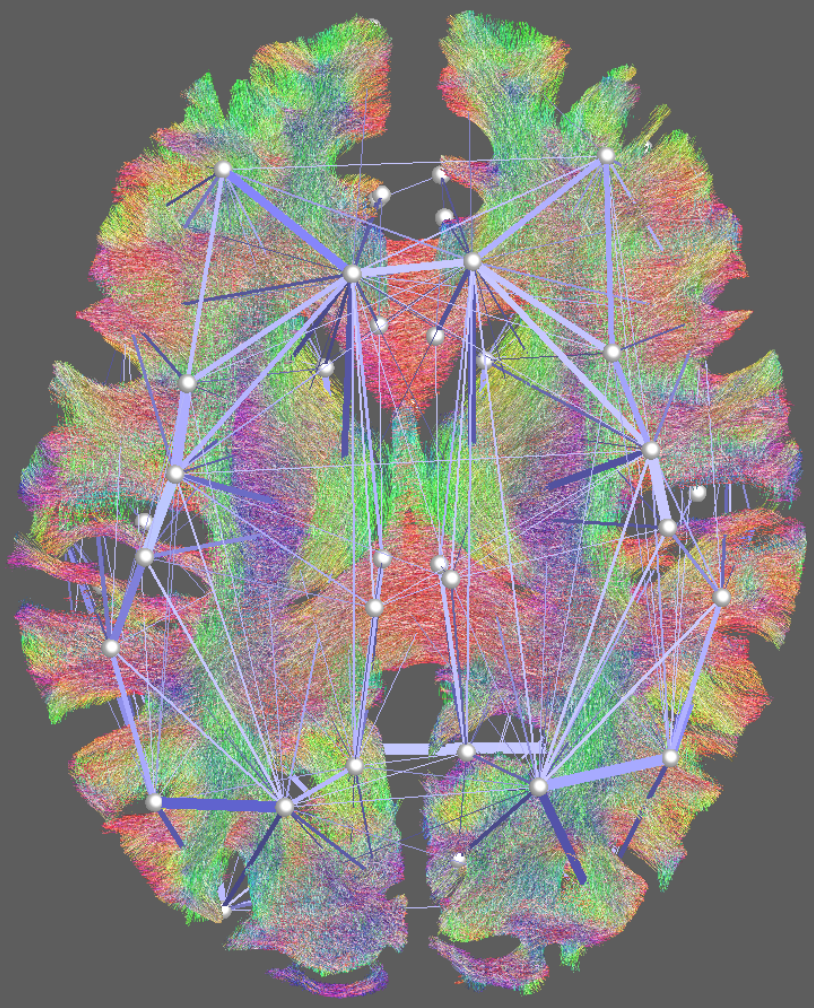
Information Processing in Neural Networks

Learning of Structural Connectivity & Dynamics of Functional Activation



Holger Ewald Finger

Information Processing in Neural Networks:

Learning of Structural Connectivity and Dynamics of Functional Activation

Dissertation
zur Erlangung des Grades
Doktor der Naturwissenschaften (Dr. rer. nat.)
im Fachbereich Humanwissenschaften der
Universität Osnabrück

vorgelegt von

Holger Ewald Finger

Osnabrück, 2016

Supervisors:

Prof. Dr. Peter König
University of Osnabrück, Osnabrück, Germany

Prof. Dr. Claus C. Hilgetag
University Medical Center Eppendorf, Hamburg, Germany

Additional reviewer:

Prof. Dr. Gordon Pipa
University of Osnabrück, Osnabrück, Germany

Curriculum Vitae

Holger Ewald Finger

Mozartstr. 3
49078 Osnabrück
Germany

Peer-Reviewed Journal Articles

2016 Holger Finger, Marlene Bönstrup, Bastian Cheng, Arnaud Messé, Claus Hilgetag, Götz Thomalla, Christian Gerloff, and Peter König (2016). "Modeling of Large-Scale Functional Brain Networks Based on Structural Connectivity from DTI: Comparison with EEG Derived Phase Coupling Networks and Evaluation of Alternative Methods along the Modeling Path". In: PLoS Comput Biol 12.8, pp. 1–28. DOI: 10.1371 / journal.pcbi.1005025.

2016 Moritz Köster, Holger Finger, Maren Kater, Christoph Schenk, and Thomas Gruber (2016). "Neuronal oscillations indicate sleep-dependent changes in the cortical memory trace". In: Journal of Cognitive Neuroscience

2016 Caspar M. Goeke, Serena Planera, Holger Finger, and Peter König (2016). "Bayesian Alternation during Tactile Augmentation". In: Frontiers in Behavioral Neuroscience 10, p. 187. ISSN: 1662-5153. DOI: 10 . 3389 / fnbeh.2016.00187. URL: http://journal.frontiersin.org/ article/10.3389/fnbeh.2016.00187

2013 Holger Finger and Peter König (2013). "Phase synchrony facilitates binding and segmentation of natural images in a coupled neural oscillator network". In: Frontiers in computational neuroscience 7.195

Peer-Reviewed Conference Articles with Talks

2011 Holger Finger and Shih-Chii Liu (2011). “Estimating the location of a sound source with a spike-timing localization algorithm”. In: 2011 IEEE International Symposium of Circuits and Systems (ISCAS). IEEE, pp. 2461– 2464

2010 Holger Finger, Shih-Chii Liu, Paul Ruvolo, and Javier R Movellan (2010). “Approaches and databases for online calibration of binaural sound localization for robotic heads”. In: Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on. IEEE, pp. 4340–4345

Other Workshop Talks

2016 Holger Finger, "From DTI to EEG: Modeling Functional Connectivity based on the Structural Connectome", In the modeling session of the retreat of the SFB 936, Boltenhagen, November 2016

2015 Holger Finger, "Functional Binding in Neural Networks". Summer School Information Processing in Neural Systems: From Single Neurons to Large-Scale Models of Cognition, Osnabrück, May 2015

Poster Contributions

2016 Marlene Bönstrup, Holger Finger, Bastian Cheng, Arnaud Messé, Claus Hilgetag, Götz Thomalla, Christian Gerloff, and Peter König (2016) Modeling of large-scale functional brain networks based on structural connectivity from DTI: comparison with EEG derived phase coupling networks and evaluation of alternative methods along the modeling path. San Diego, November 2016

2015 Moritz Köster, Holger Finger, Maren Kater, Christoph Schenk, Thomas Gruber (2015) Sleep Dependent Memory Consolidation and its Effects on Neuronal Oscillations in the Human EEG. Psychology and Brain 2015, Frankfurt am Main, June 2015

2013 Holger Finger and Peter, König (2013) Phase-Synchrony Facilitates Binding and Segmentation of Natural Images in a Neural Network Model. 10th Göttingen Meeting of the German Neuroscience Society, Göttingen, March 2013

2013 Holger Finger and Peter, König (2013) Emergence and plasticity of architectures underlying multi-site communication, SFB 936 Retreat, Boltenhagen, October 2013

2012 Holger Finger and Peter König (2012) A Normative Model of Neuronal Dynamics for Visual Feature Binding Based on Phase-Synchrony and Natural Image Statistics, Bernstein Conference, Munich September 2012

Supervised B.Sc. and M.Sc. Theses

2016 Investigating the effects of spatial resolution on structural connectivity in the brain. Master thesis of Arushi Garg.

2016 Binding in natural images through phase synchronization in complex-valued autoencoder networks. Bachelor thesis of Kai Standvoß.

2015 Estimation and Analysis of Structural Connectivity Based on the Optimization of Prediction of Functional Connectivity. Bachelor thesis of Patrick Winfried Josef Ebel.

2013 Optimization of Phase Synchrony in a Neural Network Simulating Visual Feature-binding. Bachelor thesis of Caroline Fischer.

2013 Auditory scene analysis in an oscillating neural network. Bachelor thesis of Florian Sandhäger.

Teaching

WS 2016 Deep Learning Reading Club

SS 2009 Teaching Assistant in Mathematics II

WS 2008 Teaching Assistant in Mathematics I

Abstract

Adaptability and flexibility are some of the most important human characteristics. Learning based on new experiences enables adaptation by changing the structural connectivity of the brain through plasticity mechanisms. But the human brain can also adapt to new tasks and situations in a matter of milliseconds by dynamic coordination of functional activation. To understand how this flexibility can be achieved in the computations performed by neural networks, we have to understand how the relatively fixed structural backbone interacts with the functional dynamics. In this thesis, I will analyze these interactions between the structural network connectivity and functional activations and their dynamic interactions on different levels of abstraction and spatial and temporal scales.

One of the big questions in neuroscience is how functional interactions in the brain can adapt instantly to different tasks while the brain structure remains almost static. To improve our knowledge of the neural mechanisms involved, I will first analyze how dynamics in functional brain activations can be simulated based on the structural brain connectivity obtained with diffusion tensor imaging. In particular, I will show that a dynamic model of functional connectivity in the human cortex is more predictive of empirically measured functional connectivity than a stationary model of functional dynamics. More specifically, the simulations of a coupled oscillator model predict 54% of the variance in the empirically measured EEG functional connectivity.

Hypotheses of temporal coding have been proposed for the computational role of these dynamic oscillatory interactions on fast timescales. These oscillatory interactions play a role in the dynamic coordination between brain areas as well as between cortical columns or individual cells. Here I will extend neural network models, which learn unsupervised from statistics of natural stimuli, with phase variables that allow temporal coding in distributed representations. The

analysis shows that synchronization of these phase variables provides a useful mechanism for binding of activated neurons, contextual coding, and figure ground segregation. Importantly, these results could also provide new insights for improvements of deep learning methods for machine learning tasks.

The dynamic coordination in neural networks has also large influences on behavior and cognition. In a behavioral experiment, we analyzed multisensory integration between a native and an augmented sense. The participants were blindfolded and had to estimate their rotation angle based on their native vestibular input and the augmented information. Our results show that subjects alternate in the use between these modalities, indicating that subjects dynamically coordinate the information transfer of the involved brain regions. Dynamic coordination is also highly relevant for the consolidation and retrieval of associative memories. In this regard, I investigated the beneficial effects of sleep for memory consolidation in an electroencephalography (EEG) study. Importantly, the results demonstrate that sleep leads to reduced event-related theta and gamma power in the cortical EEG during the retrieval of associative memories, which could indicate the consolidation of information from hippocampal to neocortical networks. This highlights that cognitive flexibility comprises both dynamic organization on fast timescales and structural changes on slow timescales.

Overall, the computational and empirical experiments demonstrate how the brain evolved to a system that can flexibly adapt to any situation in a matter of milliseconds. This flexibility in information processing is enabled by an effective interplay between the structure of the neural network, the functional activations, and the dynamic interactions on fast time scales.

Contents

List of Figures

List of Tables

Chapter 1

General Introduction

Humans have to adapt on a moment-by-moment basis to new situations. We can instantly recall required items from memory or utilize specialized cognitive functions. At the same time, we as human beings have the feeling that our thoughts can wander freely from one idea to another without any barrier that could impede this process. How is this seemingly barrier-free flexibility in our mind possible on a mostly static and fixed substrate?

To understand this question, one first has to realize that information in the brain is stored and processed in a distributed way. Specific items in memory are represented in localized brain areas, cell populations or even individual neurons (O'Keefe and Recce, 1993; Buzsáki and Moser, 2013; Perrett, Rolls, and Caan, 1982). Furthermore, the brain is organized in many functionally specialized areas that are necessary to execute specific functions (Mazoyer et al., 2001). Therefore, the brain has to employ the right cortical resources at the right time and eventually combine information from several areas and their resulting subprocesses to solve a given task.

The flexibility in information processing capabilities requires dynamic coordination between and within brain areas and cell assemblies (Malsburg, Phillips, and Singer, 2010). This dynamic coordination is comprised of several key aspects: (1) dynamic binding, linking or grouping of information (Gray et al., 1989; Gray and Singer, 1989; Singer, 2004; Kay and Phillips, 2011), (2) dynamic routing of information flow (Fries, 2005), and (3) the activation and deactivation of relevant brain processes (Malsburg, Phillips, and Singer, 2010). In this thesis this dynamic coordination is analyzed with computational and experimental methods on small and large spatial scales, as well as on slow

and fast time scales. The computational experiments specifically model oscillatory brain interactions using relatively simplified phase oscillators that have a regular limit cycle with a predefined frequency (Kuramoto, 1975; Moreno and Pacheco, 2004). Based on such simplified models of the dynamic process, we can tackle questions about their functional role and the computational principle they might serve. To address this issue, in one study I compare such a dynamic phase oscillator model with a model of only stationary dynamics regarding their ability to predict empirically measured functional brain connectivity based on the underlying structural brain connectivity (Finger et al., 2016). In another study, I will show how these phase oscillators can implement binding by synchrony (Milner, 1974; Malsburg, 1981; Gray and Singer, 1989) in an artificial neural network that is learned unsupervised from natural visual stimuli (Finger and König, 2013).

In addition to the cognitive flexibility that is achieved by dynamic coordination of information processing through oscillatory couplings, there is also cognitive adaptation by learning through neural plasticity mechanisms (Hebb, 1949; Song, Miller, and Abbott, 2000). It is largely assumed that these plasticity mechanisms depend mostly on the local network activations of the pre- and postsynaptic neurons, although some global signaling mechanisms might be involved as well (for a review of dopamine and learning see Wise, 2004). Therefore, the routing of information to the relevant structurally connected cells is an important prerequisite for these local learning rules. As a result, there could be two-way interactions between the dynamic coordination of functional activation and learning through changes of structural connectivity. In fact, the overall long-term process of learning consists of several of these synaptic modifications that could take place over a longer period of time to consolidate information in the structure of the network (McGaugh, 1966). Therefore, the neurophysiological mechanism of synaptic weight modification connects the dynamic network activation on fast timescales to the long-term process of learning on slow timescales (Meyer-Bäse, Ohl, and Scheich, 1996).

In order to gain new insights about the computational principles the brain implements, I will model these changes in the structural connectivity using normative models of learning (Wyss, König, and Verschure, 2006; Einhäuser and König, 2010), which aim to abstract away biological implementation details that are not necessary for the functional role of learning. These mechanisms

are closely related to the emerging field of deep learning and artificial neural networks in the machine learning community. The combination of ideas and concepts from the fields of neuroscience and deep learning bears a high potential for new discoveries. For example, new insights from neuroscience and cognitive science might be very relevant for the advancement and development of new approaches to artificial intelligence (AI) and might offer solutions to highly complex machine learning tasks such as autonomous driving or AI agents for video games (Mnih et al., 2013). But similarly, also new insights from deep learning might shed novel light on the interpretation of neuroscientific data, especially regarding the implementation of computational principles in the brain.

Learning is facilitated by the repetitive presentation of the stimuli or repetitive executions of the involved actions so that the corresponding neural activation patterns are consolidated into changes of structural connectivity (Hebb, 1949). In fact, this repetitive activation of the corresponding neural populations might not even require repeated sensory stimulation, but could even be accomplished by a working memory buffer in the hippocampus (Squire and Wixted, 2011). Lisman and Jensen argue that the hippocampal system could implement a working memory buffer by the activation of each memory item in a different cycle of the gamma rhythm within a slower theta cycle (Lisman and Jensen, 2013). Similarly, sleep can also reactivate the network through stimulation of the corresponding activation patterns. Thereby, newly acquired knowledge that was previously only stored in a short-term memory buffer with limited capacity can be consolidated (Born and Wilhelm, 2012). In this regard, we investigated the role of sleep in relation to learning of associative memory items (Köster et al., 2016).

Overall, this thesis is about the interplay between structural connectivity, functional activation, and dynamic organization on different timescales. Figure 1.1 illustrates the interactions between these different abstract concepts. This interplay is investigated from different perspectives. The experiments will address the following questions:

(A) How the relatively static structure of a neural network influences functional activations, oscillatory processes and their dynamic interactions on fast timescales (arrows from left to right in Figure 1.1).

(B) How the instantaneous network activation influences the structure of the network through learning (arrow from the center to left in Figure 1.1) and how it induces dynamic interactions on fast timescales to allow temporal coding in distributed representations (arrow to and from the right circle in Figure 1.1).

(C) How these mechanisms facilitate behavioral and cognitive functions through multisensory integration and memory consolidation and retrieval (arrow to and from the top circle in Figure 1.1).

Although a variety of experimental approaches and techniques are used throughout this thesis, the different subprojects are unified by the overall question of how relatively static structural connectivity interacts with the highly dynamic functional connectivity to allow for fast switches in cortical processing. Throughout this thesis I will highlight the different temporal scales in which these interactions take place. Specifically, as illustrated in Figure 1.1, these interactions include learning processes on slow time scales, functional activation of brain processes based on the underlying structure and behavioral perception on medium timescales, and oscillatory processes that can dynamically reorganize the effective interactions on very fast timescales. Combined, this will help us to understand how complex cognitive functions can emerge from this interplay between brain areas, cell assemblies and individual neurons and their dynamic functional interactions.

There are multiple spatial scales and levels of abstraction at which these dynamic interactions could occur. On the lowest level, individual neurons are grouped into cell assemblies in the microscale circuitry (Cajal, 1909; Martin and Whitteridge, 1984; Helmstaedter, Briggman, and Denk, 2008) and might organize their functional interactions with precise spike timings (Markram et al., 1997; Song, Miller, and Abbott, 2000). These cell assemblies in turn form cortical columns (or cortical modules, see Casanova, 2005), which might be functionally organized by synchronized coactivations of many neurons, reflected in oscillations in the local field potential (Murthy and Fetz, 1996; Goldberg et al., 2004). On the next level, brain areas are composed of these small modules (Brodmann, 1909; Sporns, Tononi, and Kötter, 2005) and might be coactivated based on the current situation or task (Aguirre, Zarahn, and d'Esposito, 1997; Horwitz, 2003; Simmonds, Pekar, and Mostofsky, 2008).

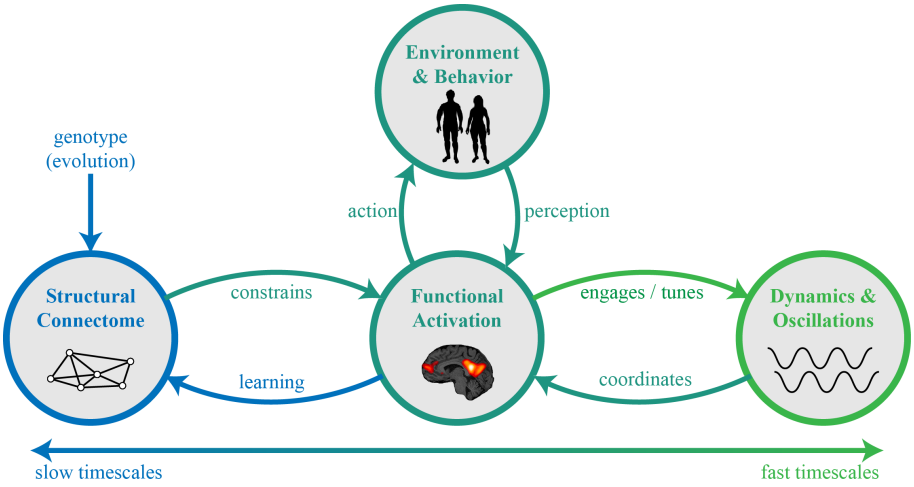


FIGURE 1.1: **Overview of interactions on different timescales:** The network structure and the perception of the environment determine the functional activations. These activations automatically induce oscillations on faster timescales. Conversely, the dynamic interactions on fast timescales help to coordinate the functional activation. Finally, the functional activation can influence the structure by learning mechanisms.

Lastly, interactions between these brain areas enable cognitive functions (Mazoyer et al., 2001; Sporns et al., 2004), such as working memory (Hampson et al., 2006), perception of multisensory concepts (Grossman et al., 2000), and complex interactions with the environment (Pelphrey, Morris, and Mccarthy, 2004).

This thesis contains four peer-reviewed publications that will be presented in the subsequent chapters. For a clearer description, in the following these studies will be named (1) connectome modeling, (2) temporal coding, (3) sensory integration, and (4) memory consolidation. These studies are organized into three chapters based on their corresponding spatial scale and level of abstraction. First, chapter 2 lays the fundamental ground work, with the connectome modeling study (Finger et al., 2016) that describes a computational model of functional interactions on the level of the full brain. This study will investigate several alternative methods along the modeling pipeline from the estimation

of structural connectivity via computational models to the estimation of functional connectivity from EEG recordings. Next, chapter 3 presents the temporal coding study (Finger and König, 2013) that focuses on the microscopic circuits and models of representation learning from natural visual stimuli. Importantly, I extend the rate-based units with phase oscillators to investigate binding by synchrony in the setting of unsupervised learning. Chapter 4 focuses on the relation to behavior and cognition and contains two studies: The first is the sensory integration study (Goeke et al., 2016), which investigates the integration of rotation cues of an augmented tactile modality and the native vestibular modality. The last study is about associative memory consolidation during sleep (Köster et al., 2016). It investigates sleep-dependent changes in cortical oscillations during the retrieval of associative memory items. Finally, the discussion chapter will continue the synopsis of my work by highlighting the overarching consequences of the presented studies.

This introduction is organized into the following six main sections: In the first section I will present the overall modeling approach that combines the interactions on slow, medium, and fast timescales. The second section will explain how this modeling approach is applied to bridge the gap between empirical measurements of structural and functional connectivity in the large-scale connectome. Conversely, the third section will introduce the reverse effect: namely, how functional activation changes the structural connectivity of networks through unsupervised representation learning. The fourth section will describe how important temporal coding and synchrony of oscillations are in these distributed representations. The fifth part of this general introduction will relate these concepts to behavioral cognitive functions, such as Bayesian cue combination, and memory consolidation during sleep. Finally, the last section will provide an overview of the manuscripts and their relation to the overarching topics of this thesis.

1.1 Modeling Approach Bridging Multiple Timescales

Structural connectivity describes the anatomical backbone of a neural network, while functional connectivity describes the instantaneous network activations and interactions that take place at any given moment in time. A descriptive analogy is that functional connectivity on a structural network is like software running on computer hardware. Similar to brain connectivity, there are different software components that might interact at different time points to solve a given task, while the hardware provides the necessary substrate for the computations to be run. However contrary to the fixed computer hardware, the structural connectivity of biological neural networks might be changed dynamically by learning mechanisms (section 1.3). Furthermore, the functional connectivity of biological neural networks shows oscillatory dynamics on very fast timescales (section 1.4). On the whole, the human brain shows more dynamic two-way interactions that bridge multiple timescales between the structure, the functional activation, and dynamic oscillatory couplings.

The categorization of cortical connectivity into either structural or functional connectivity is useful to explain their interactions. For example, changes in structural connectivity take place on slower timescales through neural plasticity mechanisms mediated by functional connectivity repeating on fast timescales (Markram et al., 1997; Bi and Poo, 1998). Conversely, functional connectivity is to a large extent influenced and constrained by the structural connections of the neural network. But in contrast to the structure, functional connectivity also flexibly adapts to the current context of the situation or to a given behavioral or cognitive task (Cordes et al., 2000; Elton and Gao, 2014). Underlying this flexible adaptation are theories that the communication in the brain is influenced by coherence (Engel, Fries, and Singer, 2001; Fries, 2005). On smaller spatial scales, similar theories exist for the inter-columnar synchronization in the visual cortex coding global stimulus attributes (Gray et al., 1989)

In this section, I will introduce the modeling approach that is used to model interactions between structural plasticity on slow timescales, functional activation on medium timescales, and dynamic oscillatory organization on fast timescales. In the following I will first introduce several modeling approaches to model dynamic interactions in cortical networks on different levels of abstraction. Afterwards, I will present a new combination of models including

phase and rate variables to simulate interactions bridging multiple timescales.

Dynamical Systems and Phase Oscillators

As mentioned in the beginning of this introduction, the ability to switch flexibly and instantly between tasks or thoughts is a very important characteristic of the human brain. From a theoretical standpoint, this could be achieved by a dynamical system if it has multiple stable equilibrium points in the vector space spanned by the states of the system (Amit, 1992; Samsonovich and McNaughton, 1997; Lu, 2002). This multistability property of a dynamical system allows abrupt phase transitions and thereby fast switching between behavioral tasks (Kelso, 2012).

This is also reflected in behavioral phenomena of multistable perception. A famous example of such a phenomena is the Necker cube (Necker, 1832), which is a transparently drawn cube so that all edges in the foreground but also in the background are visible. This cube can be perceived either as viewed from above or from below. The perceptual state is stable over time with a resistance that prevents an easy switch between the two states. The simultaneous perception of both states is usually not possible. Therefore, according to a dynamical system interpretation of perception, the perceptual state depends on the current equilibrium state (or attractor state) of the neural activations (Kawamoto and Anderson, 1985). As a result, the description of a neural network as a dynamical system can explain the dynamic organisation of functional activation based on a fixed structural backbone.

There exist many computational models that simulate neurons and their interactions as dynamical systems (Hodgkin and Huxley, 1952; Horwitz, 1990; Jansen and Rit, 1995; Izhikevich, 2003; Breakspear, Jirsa, and Deco, 2010; Woolrich and Stephan, 2013). In this section, I will just briefly introduce a few relevant models to highlight their specific differences in their level of abstraction. These differences are not only important in relation to the connectome modeling study, but also in relation to the computational principles of temporal coding, which will be introduced in more detail in section 1.4.

To begin with, most biologically motivated models of neural networks are formulated as a dynamical system that requires the integration of a set of ordinary

differential equations (ODE's) over time to simulate the timeseries of each neuron in the network. The most common numerical integration methods for differential equations are Euler's method and Runge-Kutta methods (for a review see Butcher, 1996). These numerical methods integrate the differential equations resulting in timeseries data describing the evolution of the system states over time. For example, biologically plausible neurophysiological models, such as the Hodgkin-Huxley model, can describe the physical properties of the membrane conductance and ion channels of neurons (Hodgkin and Huxley, 1952). The numerical integration of the differential equations of such neurophysiologically plausible models usually results in a dynamic limit cycle describing the oscillation. Depending on the specific parameters and neuronal inputs, the limit cycle might describe the evolution of the membrane potential, with a spike occurring at a specific point in the limit cycle. Similarly, a simulation of a whole network of neurons could result in even more complex behaviors of the dynamical system. Simplifications of such detailed models are possible by a reduction to the spike generating mechanism with simpler models of integrate-and-fire neurons (Abbott, 1999).

In contrast to these detailed neurophysiological models, neural mass models average over large populations of neurons by simulating only the mean-field approximation (Jansen and Rit, 1995; David and Friston, 2003). Figure 1.2 shows an example of a neural mass model simulating the interactions between populations in the supragranular, granular, and infragranular cortical layers (Moran et al., 2007). As illustrated at the bottom left of the figure, a simulation of the interaction between several of these neural mass models is achieved by connecting them through the structural connectome obtained from diffusion tensor imaging (DTI, see section 1.2). On the bottom right, the figure also highlights how transcranial alternating current stimulation (tACS) is implemented in the model to investigate external stimulation of the network. Overall, the corresponding differential equations simulate the changes of the average membrane potential of the neural populations (expressed by variables ν) and the average post-synaptic currents (expressed by variables i). The simplification from single cells to neural masses allows to simulate a larger area of neural tissue or even the full brain, which would be impossible to do for detailed Hodgkin-Huxley type models.

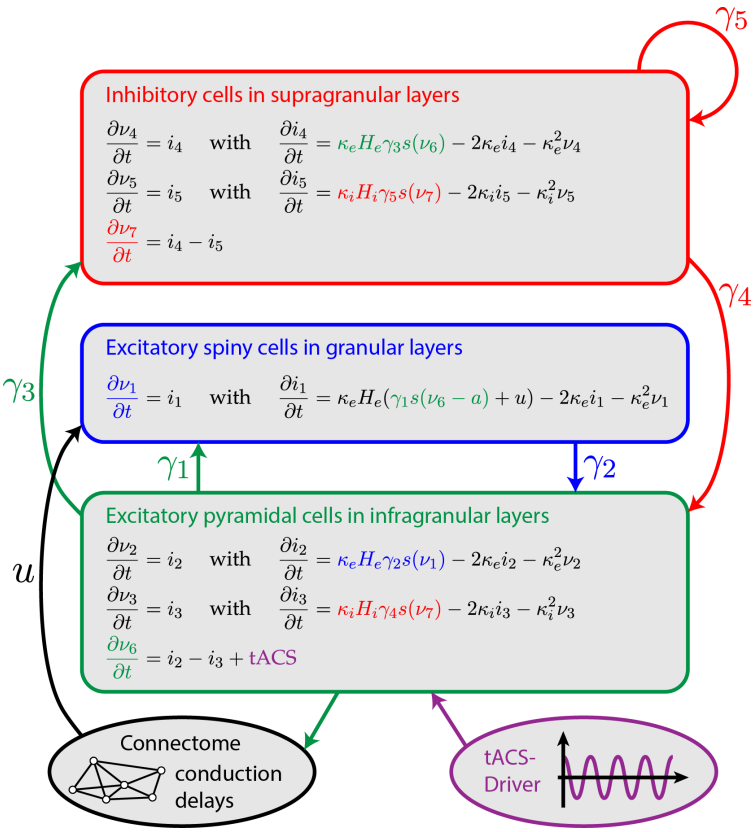


FIGURE 1.2: **Neural mass model in the connectome.** A schematic of a neural mass model showing the interactions between three different neuronal populations in the infragranular, granular, and supragranular layer. Always two differential equations (for ν and i) are used to model the postsynaptic potential for each individual connection between two populations. Here, several of these neural mass models are interacting through the structural connectome as illustrated at the bottom left. The model was also extended with an external tACS-driver that modulates the average membrane potential of excitatory pyramidal cells, as shown at the bottom right.

One might ask if further simplifications or abstractions of computational models of neural networks are possible. Neural mass models simplify the simulations by averaging over the activations of many neurons in each population. Another simplification can be achieved by completely averaging the functional activation over time (Tononi, Sporns, and Edelman, 1994). This allows the analytic calculation of the functional connectivity matrix under the assumption of stationary dynamics without simulations of the evolution of timeseries. For example, the spatial-auto-regressive (SAR) model can be solved analytically to directly obtain a functional connectivity matrix based on the structural connectome (Messé et al., 2015). This SAR model is used in the connectome modeling study (chapter 2) to evaluate the functional connectivity under the assumption of stationary dynamics.

Another alternative is to further simplify the dynamic behavior of neurons by assuming a perfect circular limit-cycle with the phase progression determined by a fixed frequency. This is an intermediate level of abstraction that specifically models the oscillatory nature of neuronal signals, but without more complex dynamics that could occur during the oscillatory limit cycle (Sturm and König, 2001). A simple framework describing network interactions between such phasor units is the Kuramoto model (Kuramoto, 1975; Breakspear, Heitmann, and Daffertshofer, 2010). This model has the advantage that it can be parameterized by only a small number of values, making it easy to explore the full potential parameter space. In fact, it describes the interactions between phasor units using a simple sinusoidal function as a phase response curve, as shown in equation 1.1.

$$\frac{\partial \varphi_j(t)}{\partial t} = \omega_j - \sum_{i \neq j} S_{ij} \cdot \sin\Big(\varphi_j(t) - \varphi_i(t - D_{ij})\Big), \tag{1.1}$$

where ω_j is the predefined frequency, with which the phase φ_j of oscillator j is progressing, and S_{ij} defines the anatomical connection strengths in the network. In comparison to the above mentioned SAR model, a network of these simple Kuramoto oscillators generates timeseries representations of the neuronal activations and even shows multistable states if time delays (D_{ij}) between network nodes are present (Kim, Park, and Ryu, 1997). Albeit this model is very simple, it allows to simulate dynamic synchronizations of neuronal assemblies. Therefore, I use variants of this Kuramoto model not only in

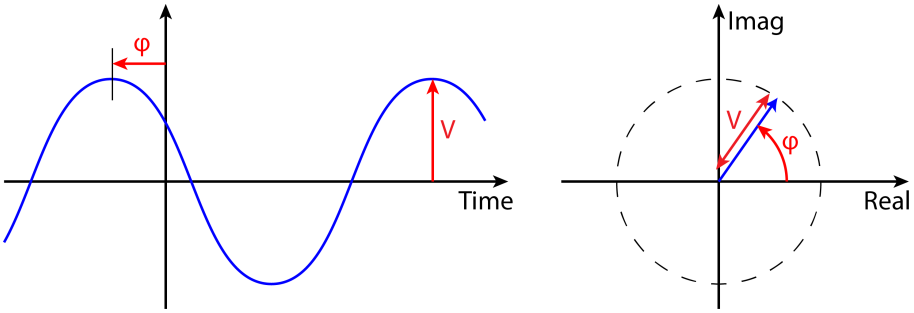


FIGURE 1.3: **Phasor model of oscillations.** The timeseries representation of the oscillation on the left is defined by the phase φ and amplitude V. This phase φ and amplitude V can be represented in the complex plane as the angle and absolute value of a complex number.

the connectome modeling study (chapter 2) to simulate functional connectivity at the full brain level (Finger et al., 2016), but also in the temporal coding study (chapter 3) to simulate binding by synchrony (Milner, 1974; Malsburg, 1981; Gray and Singer, 1989) within the visual subsystem (Finger and König, 2013).

Combining Activation and Phase in a Normative Model

The definition of network interactions in the Kuramoto model is based on the relative phase between the pre- and the postsynaptic neurons. In this standard definition of the Kuramoto model the oscillators are modeled only by their corresponding phase while the amplitude (or activation) of these oscillators is fixed. To combine activations and oscillatory phases in a normative model, the idea is that the limit cycle of a neural oscillator can be represented by a simplified phasor model as illustrated in Figure 1.3 (Fox and Bolton, 2002). This phasor model simplifies the neuronal oscillation cycle under the assumption that each neuron has a prespecified frequency, so that the oscillating timeseries $y(t) = V \cdot cos(\omega t + \varphi)$ is represented by the analytic representation $y(t) = \Re\left\{V \cdot e^{i(\omega t+\varphi)}\right\}$ as a complex number. In this representation, the phase of the oscillatory signal is coded as the complex argument (angle) and the activation as the absolute value.

With the phasor model of complex valued units in mind, complex-valued activation functions can be defined to model the influence of the presynaptic phase variables on the activation of the postsynaptic neuron. Reichert and Serre defined such an activation function in a neurophysiologically plausible way, which amplifies the influence of presynaptic neurons if they have similar phases (Reichert and Serre, 2013). More precisely, if a subset of the presynaptic neurons have a common phase, they would predominantly drive the output of the postsynaptic neuron while all remaining inputs with other phases would have an attenuated effect. Hence, this activation function flexibly adapts the routing of information through the network depending on the relative phases between the nodes.

To define phase interactions between units that have an activation determined by the receptive field inputs, Sompolinsky and colleagues extended the Kuramoto model with activation variables V_i, which modulate the phase interactions between the network nodes as follows:

$$\frac{\partial \varphi_j(t)}{\partial t} = \omega_j + \eta_j(t) - \sum_{i \neq j} V_i V_j S_{ij} \cdot \sin\Big(\varphi_j(t) - \varphi_i(t)\Big), \qquad (1.2)$$

where $\eta_j(t)$ is white noise (Sompolinsky, Golomb, and Kleinfeld, 1990). They used this model to show that dynamic grouping of neurons is possible with binding by synchrony in distributed representations (Sompolinsky, Golomb, and Kleinfeld, 1990). To model these effects, Sompolinsky and colleagues used hardcoded values for the activation levels V_j of the units and for the coupling connections S_{ij} between them. In contrast, in the temporal coding study (chapter 3) we use normative models of representation learning from natural images to determine the activation values V_j as well as the coupling strengths S_{ij} (Finger and König, 2013). Further details of the simulations and results will be provided in section 1.4.

The Kuramoto model of coupled ocillators can be simplified further under the assumption that the frequency of all nodes in the network is the same. Based on this assumption, in our study we use a transformation to a new variable that represents only the relative phase in relation to a global oscillation rhythm (Finger and König, 2013). Then the progression of the oscillatory phase does not need to be simulated, but only relative phase changes between network units. Consequently, the global oscillation rhythm, which is not simulated but

only implicitly modeled, could then be interpreted as the oscillation of a local-field potential or overall gamma rhythm of a brain area.

In summary, this section introduced several neural network models that simulate dynamic network interactions on different levels of abstraction. Starting from detailed Hodgkin–Huxley and integrate-and-fire-neurons, I introduced simplifications resulting in neural mass models, and spatial-autoregressive (SAR) models that can be solved analytically. Importantly, I showed a further simplification of the oscillatory limit-cycle to Kuramoto models. In the following section, the performance in predicting large scale functional brain connectivity of the SAR model of only stationary dynamics will be compared to a Kuramoto model with simplified oscillatory couplings with conductance delays (Finger et al., 2016). Section 1.3 will describe methods of representation learning that are used in the temporal coding study to generate the activation levels V_j for the above mentioned coupled oscillator interactions. Finally, this will be used in section 1.4 where I will focus on the functional role of such coupled oscillators for the dynamic organization of network interactions. There I will also present the simulation results of a combined model of representation learning with rate-based units extended with phase variables for binding by synchrony (Finger and König, 2013).

1.2 Functional Interactions in the Human Structural Connectome

The previous section introduced the overall modeling approaches of this thesis. Based on these approaches, the connectome modeling study (chapter 2) bridges all steps that are necessary to simulate large-scale functional EEG connectivity based on the anatomical structural connectivity (Finger et al., 2016). The motivation of this study is based on findings of Hagman and colleagues that there is a strong correspondence between the structural connectivity and functional resting-state connectivity in functional magnetic resonance imaging (fMRI) measurements from human subjects (Hagmann et al., 2008). Based on these findings, several studies used computational modeling approaches based on structural connectivity to simulate functional connectivity of fMRI recordings (Honey et al., 2009; Cabral et al., 2011; Messé et al., 2014). More recently, Cabral and colleagues used similar approaches to also simulate functional connectivity in slowly changing power envelopes (<1 Hz) of faster EEG signals (Cabral, Kringelbach, and Deco, 2014; Cabral et al., 2014). In the connectome modeling study (chapter 2) we use similar computational approaches to bridge this gap from structural to functional data, but focus on the functional connectivity of faster oscillatory dynamics obtained from EEG measurements and not only the envelope signals (Finger et al., 2016). Along the way, the study evaluates alternative processing methods not only in the computational modeling approach but also in the processing of the structural and functional data. This evaluation is made possible by using the predictive performance of the model to bridge between the structural and the functional modalities.

The following two subsections will shortly introduce the empirical measurement techniques and processing methods that are used to construct the structural and functional connectivity matrices of human subjects. This will be followed by a description of the difference between structural and functional connectivity, and approaches to bridge these differences with computational models. Finally, I will describe the results of the computational experiments and relate these to possible applications.

Structural Connectivity

Structural connectivity on the full brain level can be obtained by diffusion tensor imaging (DTI) and subsequent fiber tractography algorithms that follow fiber orientations through the white matter of the brain (Basser et al., 2000, review by Mori and Zijl, 2002). Other methods of obtaining more fine grained structural connectivity would be neuron tracing methods, which use either light microscopy or electron microscopy images (review by Lu, 2011). In comparison to these other methods, diffusion tensor imaging has the advantage that it can be used non-invasively, in-vivo and globally for the reconstruction of anatomical connections of the full brain. Therefore, the structural connectivity data in the connectome modeling study is based on diffusion tensor imaging and a subsequent fiber tractography algorithm (Finger et al., 2016).

The raw DTI measurements result in diffusion tensors at each brain voxel, describing the local diffusivity of water molecules in the neuronal tissue in all 3D directions (Figure 1.4A). From these tensors a fractional anisotropy index can be calculated that describes the amount of directionality at each brain voxel (for a review see Assaf and Pasternak, 2008 and Neil et al., 2002). As fractional anisotropy is usually low for voxels containing gray matter and high for voxels containing white matter, fiber tractography is constrained to white matter regions and can only reconstruct inter-cortical connections from one gray matter region to another, but cannot reconstruct the local connectivity within the gray matter.

Fiber tracking algorithms generate meaningful streamlines from the diffusion tensors by following the tensor directions from voxel to voxel starting from a seed region until reaching a specified target region (Mori and Zijl, 2002). For the experiments in this thesis, the seed and target regions are gray matter regions, defined with the freesurfer toolbox (Reuter et al., 2012). In general, there are two main subclasses of fiber tracking methods (Descoteaux et al., 2009): The simpler method is called deterministic fiber tractography and follows the preferred direction at each voxel. In contrast, probabilistic fiber tractography samples many streamlines for each seed voxel (usually around 10000 streamlines per seed voxel) and follows a direction that is sampled probabilistically from the local diffusion tensors along the streamline path. These

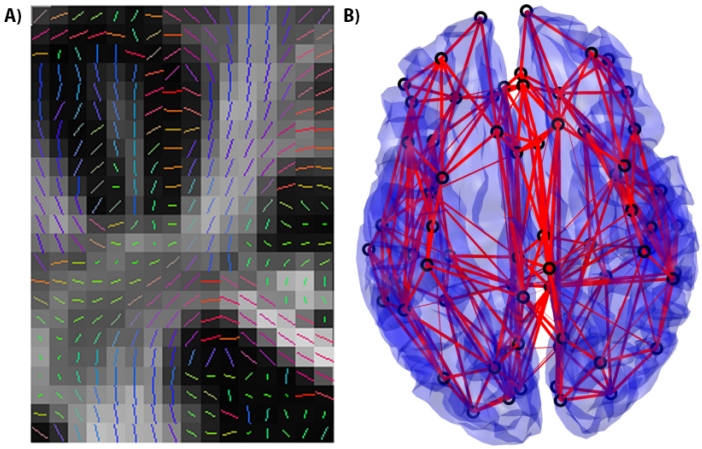


FIGURE 1.4: **Diffusion tensor imaging and structural connectivity.** (A) The panel shows a coronal view of the V1 eigenvectors as colored lines and the fractional anisotropy index per voxel as grayscale background. (B) The resulting connectivity graph of nodes corresponding to the freesurfer parcellation (superior view).

probabilistic approaches are assumed to be superior because deterministic tractography cannot model the uncertainty of the measurements and the distribution of fiber directions within each voxel (Tournier, Calamante, and Connelly, 2012). Therefore, in the connectome modeling study (chapter 2) probabilistic fiber tractography is used to obtain structural connectivity matrices from participants (see appendix A for further details).

A structural connectivity matrix from several individuals can be combined to obtain a better estimate of the wiring diagram of an average brain of a population. In this context of mapping out the full connectivity, the resulting structural connectivity is often called "connectome" (Sporns, Tononi, and Kötter, 2005). Figure 1.4B shows a graphical visualization of the resulting connectome graph with edge weights corresponding to the number of probabilistically tracked fibers (normalized and log transformed).

There are several reasons why the estimation of structural connectivity obtained from diffusion tensor imaging might not reflect the real anatomical connectivity (Kinoshita et al., 2005). For example, contradicting arguments were

presented that the resulting probabilistic connectivities underestimate (Behrens et al., 2007) or overestimate (Hagmann et al., 2008) long-range fibers. Therefore, corresponding corrective scaling terms have been suggested to correct for connection biases induced by different fiber lengths. Furthermore, intracortical tangential (within gray matter) connectivity is completely missing from the structural connectivity data obtained from white matter tractography. Therefore, some studies added additional localized intracortical connections to their models (Jirsa et al., 2010; Ritter et al., 2013).

Most fibers connecting the two hemispheres run through the corpus callosum (Schmahmann and Pandya, 2009). Many of these fibers are connecting the homologous regions of the two hemispheres and are therefore called homotopic connections. In this regard, Li and colleagues compared transcallosal fiber tractography with primate tracer studies and found that structural connectivity obtained from diffusion tensor imaging underestimates the dense homotopic connectivity (Li et al., 2012a). This was further highlighted by modeling studies of Messé and colleagues who showed that funcational fMRI connectivity is better modeled if additional homotopic connectivity is added to the structural connectivity (Messé et al., 2014; Messe, Benali, and Marrelec, 2015). Furthermore, the low resolution of the estimated structural connectivity from DTI might miss some details of the real structural connectivity. Last but not least, the diffusion tensors cannot provide information about the directionality of the connections, which could potentially have large effects on the functional interactions in the network. All these reasons motivate a comparative analysis of different preprocessing steps for the structural connectivity data. To overcome these shortcomings, in the connectome modeling study (chapter 2) different alternative methods of converting the fiber tractography data into a structural connectivity matrix are compared.

Functional Connectivity

Functional brain connectivity can be measured empirically and non-invasively with brain imaging techniques such as EEG, MEG, PET, and fMRI (Horwitz, 2003). The fMRI blood-oxygen-level dependent (BOLD) signal and the PET signals (blood flow, oxygen, glucose metabolism) do not capture the high frequency and dynamics on faster timescales that we are interested in. Therefore,

in the experiments described in this thesis we analyzed functional connectivity obtained from EEG signals. Consequently, in this section I will focus on the processing methods of these fast signals and how functional connectivity can be estimated based on these measurements.

Overall, EEG measurements result in timeseries recordings at the sensor level. Using source reconstruction methods (i.e. inverse solutions), these timeseries can be projected back to the original neural sources in specific brain areas (for a review of source reconstruction methods see Grech et al., 2008). In the connectome modeling study (chapter 2), several alternative source reconstruction methods are compared (Finger et al., 2016). In particular, we compare the LCMV beamformer (linearly constrained minimum variance beamformer, see Van Veen et al., 1997) with ELORETA (exact low resolution brain electromagnetic tomography, see Pascual-Marqui et al., 2011) and MNE (minimum-norm estimate, see Dale et al., 2000).

The estimation of functional connectivity between two nodes based on their activation is a difficult problem and an area of active research (Stam, Nolte, and Daffertshofer, 2007; Vinck et al., 2011). Most methods estimate the connectivity separately for different frequency bands. Therefore, the signals are first bandpass filtered before a Hilbert transformation extracts the phase of the signal. Finally, functional connectivity metrics like phase-locking-values (Lachaux et al., 1999) or coherence (Andrew and Pfurtscheller, 1996) can be obtained by comparing the phase and amplitudes of the two signals of network nodes. These functional connectivity metrics are not only used in the connectome modeling study (chapter 2) but also within learned neural representations of visual stimuli in the temporal coding study (chapter 3).

An important decision for the estimation of functional connectivity is the treatment of zero-phase-lag interactions between nodes. Why are zero-phase-lag interactions special in cortical dynamics? In fact, this question has a technical as well as a physiological side. On the technical side, it is clear that functional connectivity metrics, which include zero-phase-lag interactions, could include spurious correlations, resulting from volume conduction effects that do not reflect real physiological activity at the source level (Nunez et al., 1997). But on

the physiological side, there is also evidence that synchronizations with zero-phase-lag might have a special role in the brain. For example, zero-phase-interactions might occur in cortical dynamics due to mostly bidirectional connections between neurons or brain areas (Ao, Hänggi, and Schmid, 2013). A bidirectional coupling between two network nodes often results in a stable state at a phase lag of 0 (in-phase) and π (anti-phase) due to the symmetric nature of the interaction. Therefore, it might be important to retain the zero-phase-lag interactions that are of physiological nature and do not stem from volume conduction artifacts. To solve this problem, many researchers have investigated alternative functional connectivity metrics, such as imaginary coherence, phase-lag-index, lagged-phase-coherence, and a few others with special weightings of the zero-lag-contribution in the signals (Andrew and Pfurtscheller, 1996; Lachaux et al., 1999; Nolte et al., 2004; Stam, Nolte, and Daffertshofer, 2007; Vinck et al., 2011; Pascual-Marqui, 2007b). Furthermore, one has to realize that the source reconstruction methods described above already remove most correlations from the sensor signals. This raises the question whether functional connectivity metrics with removed zero-phase-lag contributions are even necessary or useful. Therefore, in the connectome modeling study (chapter 2) different metrics with and without zero-phase-lags are compared.

The nature of functional brain data is that it represents just a specific activation pattern that was present during the recording session. This activation highly depends on the task the subject is performing during the recording (Aguirre, Zarahn, and d'Esposito, 1997; Simmonds, Pekar, and Mostofsky, 2008), while the estimation of the structural connectivity is independent of any task. Therefore, I compare the simulated functional connectivity only with empirical resting-state recordings without a task context to decrease the dependence of functional connectivity on any given task. Obviously, the resulting data mainly shows the connections in the default-mode-network (DMN) and other resting-state specific networks of the brain.

Bridging the Gap Between Structural and Functional Connectivity

The empirically recorded data can be processed, as described above, with the same cortical parcellation scheme in the structural and the functional domain,

resulting in two equally sized connectivity matrices. This allows further analysis of the differences between the structural and the functional connectivity matrices as well as a computational model that can predict one from the other. Hagmann and colleagues found a substantial correspondence between structural connectivity and resting-state functional connectivity in the slowly (<1 Hz) changing signals of fMRI recordings (Hagmann et al., 2008). In fact, many following studies analyzed these relations between structural and functional data by a comparison between DTI measurements and functional connectivity obtained from fMRI recordings (Skudlarski et al., 2008; Greicius et al., 2009; Honey et al., 2009). In regard to much faster EEG functional connectivity, He and Raichle discuss evidence that the envelope signal (power) of faster EEG oscillations is related to the "slow cortical potential" as measured by fMRI recordings (He and Raichle, 2009). This was further analysed by Ko and colleagues, who found that the default mode network (DMN) as measured in the BOLD signal (Raichle et al., 2001) is related to the coherence of the slowly changing envelope signals (<0.1 Hz) of the gamma frequencies (Ko et al., 2011). Similarly, Cabral and colleagues analyzed this relationship specifically with a computational model by evaluating power envelope correlations of higher frequency oscillations (Cabral, Kringelbach, and Deco, 2014; Cabral et al., 2014). Importantly, as argued by Horwitz, there is a large difference in spatial and temporal scales between functional connectivity and effective connectivity obtained from either EEG, MEG, fMRI, or PET (Horwitz, 2003). In fact, they argue that "it is not at all clear that strong coherence in the gamma band, let us say, will correspond to strong functional connectivity in fMRI" (Horwitz, 2003). In this regard, a model of functional connectivity of fast EEG signals, could reveal further insights into the dynamic (>1 Hz) organization of effective brain connectivity, which might not be possible with predictions of slowly varying brain connectivities like BOLD signals or low-pass filtered envelope correlations. Thus, in the connectome modeling study we specifically focus on the simulation and prediction of empirically measured coherence of fast (>1 Hz) EEG signals (Finger et al., 2016).

Even if both the structural connectivity and the functional connectivity could be measured with perfect accuracy, the resulting connectivity matrices of the network nodes would be different (Sporns et al., 2004), because the former describes only the static network while the latter describes instantaneous activations. In the field of functional brain imaging, techniques such as Structural

Equation Modeling (SEM, Mclntosh and Gonzalez-Lima, 1994) and Dynamic Causal Modeling (DCM, Friston, Harrison, and Penny, 2003) aim to infer the effective connectivity in a network based on recordings of functional brain activations (for a review and comparison of functional and effective connectivity see Friston, 2011). With newly available non-invasive structural imaging techniques, a different approach is possible that models functional interactions based on the underlying structural data (Honey et al., 2009). In this regard, we want to understand what the underlying causes for the differences between structural and functional connectivity are and whether it is possible to bridge these differences using computational models of neural dynamics (Horwitz, 1990; Breakspear, Heitmann, and Daffertshofer, 2010; Breakspear, Jirsa, and Deco, 2010; Woolrich and Stephan, 2013). There are mainly two effects that could account for differences between structural and functional connectivity:

1. There might be indirect functional influences from one region to another, which are not even anatomically directly connected, but only routed via intermediate nodes.

2. The dynamics and conduction delays might lead to highly non-linear effects that could further detach the functional connectivity from the underlying structure.

Neuronal models might be able to bridge this gap between structural and functional data and could thereby reveal further insights into their interactions. Specifically, the first effect could be bridged by a very simple model that accounts only for stationary dynamics (such as the SAR model, see Tononi, Sporns, and Edelman, 1994), while the second effect might require more complex neuronal models that incorporate delayed interactions (such as the Kuramoto model, see Kuramoto, 1975). Therefore, I investigate the question, which of these computational model types can better predict the functional connectivity (Finger et al., 2016).

The connectome modeling study in chapter 2 compares multiple alternatives along the modeling path from measurements of structural brain data (DTI) to measurements of functional brain activations (EEG). To evaluate the performance of these alternative methods, the correlation between the simulated and the empirically measured functional connectivity was calculated (Finger et al., 2016). Optimally, this performance comparison of alternative methods should

be based on equal assumptions and equal parameters. However, a specific setting of parameters cannot easily be justified for any of the involved methods. Hence, to obtain an unbiased performance comparison, we decided to optimize the parameters such that the individual performance of each compared alternative method is maximized. In fact, this maximization requires a full exploration of the parameter space, which is not feasible with complex models with many parameters due to the computation time. Consequently, we limited the comparison to only the simple spatial-autoregressive (SAR, see Tononi, Sporns, and Edelman, 1994) model and the Kuramoto model (Kuramoto, 1975). We used the spatial-autoregressive (SAR) model as a reference model to calculate functional from structural data, because it is analytically solvable and depends only on one parameter (global scaling of structural connectivity) and therefore easily allows a grid search of the complete parameter space.

The study shows how the gap from empirical structural data to empirical functional data can be bridged using a computational model with different alternative methods along the modeling path (Finger et al., 2016). Specifically, the SAR model, which simulates only stationary dynamics, was able to explain 45.4% of the variance in the empirical functional connectivity. In comparison, the Kuramoto model, which includes conduction delays and dynamic interactions, was able to increase this match to 54% of variance that could be explained, which is to my knowledge the best result obtained so far for such a model of functional connectivity based on the structural connectome.

The study also investigated many alternative methods along the modeling path. For example, the results show that the addition of homotopic structural connections was beneficial for the model performance, while other alternative normalization methods of the structural connectivity had only minor effects on the performance as long as a normalization is performed that accounts for the different sizes of brain regions. Moreover, the results show that functional connectivity metrics like coherence and phase-locking-values lead to a much better performance in comparison to metrics that exclude the zero-lag interactions. The model performance is robust against exchanging the reference source reconstruction method (LCMV beamformer, Van Veen et al., 1997) with either ELORETA (Pascual-Marqui et al., 2011) or MNE (Dale et al., 2000). Finally, a comparison with a forward projection shows that the source reconstruction in itself is important to avoid confounding the results by mixing of the source timeseries in the sensor space.

To gain further insights into the strengths and weaknesses of the computational model and the corresponding methods that reconstruct the structural and functional connectivities from human subjects, we analyzed the prediction performance also locally for individual brain regions. In fact, there might be specific brain regions that are more easy or difficult to model or that are over- or underestimated by the modeled functional connectivity, depending on specific characteristics of these brain areas. In this regard, I evaluate the dependence of this local model error on several metrics. We show, that the model performance depends on specific features of the Euclidean space, such as distance between regions and size of the regions, and features of the network graph topology, such as betweenness-centrality (Freeman, 1977). We found that the model performance is lower for central hub nodes, indicating that the SAR model of only stationary dynamics might not reproduce dynamic switches of effective connectivity at these nodes. The results regarding region sizes revealed that the functional connectivity of larger regions was better estimated, which could be attributed to the limited resolution of the measurements of structural as well as functional connectivity. In this regard, I will discuss steps to improve the parcellation method in the general discussion (chapter 5).

Overall, the results of the connectome modeling study bring us closer to potential applications that might benefit from a computational model of functional connectivity based on empirically measured structural brain connectivity. As an outlook into the future, such simulations of large scale brain dynamics based on anatomical connectivity could be applied for the clinical treatment of patients with brain diseases (Falcon et al., 2015; Jirsa et al., 2016). For such clinical application, individualized simulations of brain dynamics could reveal pathological changes in the underlying physiological parameters of the network. In fact, these simulations could possibly indicate specific medical treatments to correct for deficiencies in global brain dynamics. In this regard, the connectome study (Finger et al., 2016) is an important milestone because fast oscillatory dynamics might be very relevant for cognitive functions as already mentioned in the beginning of this introduction chapter and will be highlighted further in the next sections. For example, simulations of functional connectivity of stroke patients showed the disruption of network dynamics after stroke and their relation to specific changes in global and local network parameters (Falcon et al., 2015; Falcon et al., 2016). Also, applications to epileptic seizure modeling demonstrated how to develop novel personalized strategies towards

therapy and intervention (Jirsa et al., 2016). This shows the large potential of simulating individual brain dynamics. To further improve such treatment approaches, a better understanding of these brain dynamics and their functional role is necessary. Altogether, the results of the connectome modeling study (chapter 2) help to further improve our knowledge of the mechanisms that influence the functional connectivity in the brain. In this section the investigation methods were on a descriptive level enabling the comparison with empirical measurements of brain connectivity. In contrast, the next sections will further investigate these dynamic brain processes in regard to their underlying computational principles and functional roles.

1.3 Unsupervised Learning of Structural Connectivity

The previous section compared empirical structural and functional brain connectivity to draw conclusions about the relationship between structure and function of the brain. Conversely, this section follows a more bottom-up driven approach on structural brain connectivity by focusing on the question how the structure develops and is guided by learning from the statistics of functional activation (Olshausen, 1996). Importantly, these functional activations are based on natural sensory stimuli so that the training of the neural network with unsupervised learning methods could lead to meaningful representations of these inputs (Olshausen and Field, 2004; Hinton, 2007; Bengio, Courville, and Vincent, 2013).

Different types of sensory receptor neurons are responsive to changes of environmental properties. For example, photoreceptors are sensitive to light (Baylor, 1987) and hair cells are sensitive to sound waves (Brownell et al., 1985). Learning, in the sense of neural plasticity, requires that the environmental properties are reflected in some form in a neural activation pattern, so that this activation can then change the structural connectivity between individual neurons by synaptic plasticity mechanisms (Hebb, 1949). This poses the important question, how learning of complex or abstract features is achieved only by the adaptation of individual synaptic weights in the network. This question is particularly interesting considering that these synapses are only influenced by the pre- and postsynaptic neurons without distant effects that could transmit a global learning signal. While the learning rules in this section are not implementing specific neurophysiological mechanisms and might make use of distant effects, I will deal with this question in the general discussion (chapter 5), where I will motivate their implementation with spike-timing dependent plasticity.

In contrast, here I ask the question, what the computational principles of neurons and their interactions are, while neglecting the specific neurophysiological implementation details (Hinton and Sejnowski, 1999). This question also relates to the computational principle of dynamic organization of functional network interactions, which will be described in detail in the next section (1.4),

where a computational model of binding by synchrony is implemented without explicit integration of an oscillatory limit cycle. In this regard, the temporal coding study (chapter 3) uses normative models to describe not only the learning of representations from sensory inputs, but also dynamic binding by synchrony within these representations (Finger and König, 2013). The idea of such normative models is to define the computational principle as a goal function to describe how properties of the structural network or their functional activations shall be optimized (Einhäuser and König, 2010). In fact, the unsupervised learning from the statistics in sensory inputs provides the basis for higher levels of learning, because more abstract concepts would be based on these lower level representations.

The optic nerve contains between 770 000 and 1.7 million nerve fibers through which visual information is transmitted to the primary visual cortex (Jonas et al., 1992). Obviously, the recognition of a visual object within this high bandwidth and high dimensional representation requires an enormous amount of information compression. In this regard, a normative approach of sensory learning can describe how such compressed representations are developed in the brain (Baldi, 2012). Specifically, the statistical dependencies in the high dimensional space of sensory inputs can be used by unsupervised learning methods to transform these sensory inputs into more meaningful representations (for a review see Bengio, Courville, and Vincent, 2013). Interestingly, such methods can learn from sensory statistics and thereby produce neural connection weights that resemble receptive field properties as found in the primary visual cortex (Olshausen, 1996; Einhaäuser et al., 2002; Masquelier et al., 2007).

In this section, I will first compare these normative models of representation learning in nervous systems with artificial neural networks for machine learning. Then, I will describe the biologically motivated approach to understand the computational principles underlying the canonical microcircuits of the cortex. Afterwards, a description follows how hierarchies of visual features can be learned from natural image statistics. Next, an introduction into autoencoder learning is provided, because these are used in the temporal coding study (Finger and König, 2013). Finally, I will discuss how perceptual timescales can be used in such models of unsupervised learning from sensory inputs, which further relates these models to other topics of this thesis.

Comparison to Artificial Neural Networks

Normative models of representation learning from sensory statistics usually use back-propagation of error gradients (Olshausen and Field, 1997; Hyvärinen and Hoyer, 2000; Wyss, König, and Verschure, 2006) similar to their use in deep learning models for machine learning tasks (LeCun et al., 1989; Bengio, 2009; Schmidhuber, 2015). Therefore, a comparison between these biologically inspired neural networks and artificial neural networks for machine learning seems relevant. In fact, a comparison of these applications could also advance both lines of research. For example, artificial neural networks could reveal new insights into the computational principles of the cortex, because these networks solve a specific task, which might be solved by the human brain in a similar way. But also vice-versa, insights into the biological mechanisms of the human brain might help to improve artificial neural networks for machine learning applications.

Biological networks are usually highly recurrent in comparison to simple feedforward architectures that are often used for classification tasks in machine learning. However, in recent years more and more networks for machine learning tasks also use recurrent architectures (Mikolov et al., 2010; Graves, Wayne, and Danihelka, 2014; Gregor et al., 2015; Donahue et al., 2015; Oord, Kalchbrenner, and Kavukcuoglu, 2016; Graves et al., 2016). Therefore, a comparison might be especially relevant today more than ever before. Importantly, both artificial neural networks and biological neural networks store memory in a distributed way across the network. Furthermore, in both cases learning is implemented by the adaptation of synaptic weights. Hence, one might speculate that these similarities make both types of networks more adaptable and flexible for a wide range of tasks. Taking this thought one step further, I speculate that these mechanisms of dynamic organization by oscillatory couplings might also be relevant for deep learning if they serve a specific functional role in the brain.

Flexibility and adaptability become more important with increasingly complex tasks that machine learning algorithms have to solve (Vilalta and Drissi, 2002). This flexible adaptation of already trained neural networks is now more relevant than ever in the area of deep learning. The reason is that many advanced models (often 50 layers deep), which are already trained on huge datasets,

are freely available and can be used to initialize networks for new tasks (Simonyan and Zisserman, 2014; He et al., 2015; Szegedy et al., 2015). This allows to make use of previously learned good feature representations of visual or auditory signals (Pan and Yang, 2010; Bengio, 2012). Notably, in current computer vision competitions the best performing models are often based on networks that were already extensively pretrained on large datasets of images beforehand. In this regard, usually the top most layers of the network are chopped off, so that the lower layers can be reused in a new network and in a new task setting. This transfer of an already trained network to a new task is called "transfer learning". Interestingly, Raina and colleagues show that transfer learning even from unlabeled data improves the performance in an image classification task (Raina et al., 2007), while this unlabeled data is very unspecific and does not follow the same generative distribution.

This reuse of network weights from one task to another shows a high similarity to how the human brain uses preexisting capacities to solve new tasks. Improvements in dynamic adaptation of network architectures to new tasks could probably further reduce the amount of required feature engineering for machine learning applications. Therefore, flexibility is an important feature, which will enable artificial neural networks to solve a larger variety of tasks. In this regard, the temporal coding study (Finger and König, 2013) is of interest for improvements in flexibility that might be possible due to dynamic reorganization of the functional interactions in a network of rate- and phase-coding neurons.

Canonical Microcircuits and Gradient Optimization

The sensory areas in the brain can extract complex features from the high bandwidth of sensory inputs. If we understand how these feature representations on lower sensory areas are learned, we might speculate that these principles are similar in higher cortical areas. For example, some theories suggest that cortical modules are organized in a canonical microcircuit that implements the same computational principle throughout the neocortex (Douglas, Martin, and Whitteridge, 1989; Bastos et al., 2012). If such a canonical microcircuit exists, we want to understand the general principle of computation in these microcircuits while neglecting implementation details where possible. A goal function can

be used to formulate such computational principles that should be optimized, so that these can be tested in computational experiments (Barlow, 1959; Hinton and Sejnowski, 1999; Calvo and Gomila, 2008; Einhäuser and König, 2010).

In machine learning applications, the goal is usually a specified prediction output. This is achieved by specifying either logistic output units for binary decisions, or softmax output units for multi-class-classification, or linear units for the regression of continuous values. A loss function can then be defined to minimize the difference between the network output and the target prediction (review by Dreiseitl and Ohno-Machado, 2002; Zhang, 2000, also see Bishop, 2006). In contrast, when we aim to understand the computational principles of the brain, these goal functions usually specify some property of the local activations within the network (Hyvärinen and Hoyer, 2000; Simoncelli and Olshausen, 2001; Einhaäuser et al., 2002; Wyss, König, and Verschure, 2006). This allows to apply gradient descent optimization in unsupervised settings, where the network can learn solely from the presentation of sensory stimuli.

These goal functions can then be optimized using gradient descent by iteratively changing the neural weights in such a way that the global goal function is maximized. There are multiple variations of gradient descent optimization that I will shortly introduce in the following. Usually, some form of momentum is used to speed up the convergence of the optimization by dampening oscillations in the changes of the synaptic weights (Sutskever et al., 2013). Furthermore, a stochastic version of gradient descent can improve the computational efficiency by feeding only mini-batches into the network instead of the full training set, so that weights can be updated more often (Spall, 2005). Interestingly, this type of stochastic optimization is more close to biological learning, where each perceived sample of the environment would eventually directly influence the synaptic weights. In fact, the same kind of online learning can be implemented with artificial neural networks by reducing the size of the minibatch to just one input sample. For a recent review of gradient optimization methods see Ruder (2016).

It was often argued that there could be many local minima in the energy function and that the optimization might get stuck in one that is not the global optimum (Hinton and Salakhutdinov, 2006). However, more recently it has been shown that this is usually not a problem in high dimensional parameter spaces like neural network weights (Dauphin et al., 2014). In fact, Dauphin

and colleagues show that in higher dimensional spaces the propability of local minima with a high loss is drastically decreased. A more pressing problem are saddle points that could slow down the learning process. Therefore, it is important to use optimizers that can easily escape from these regions with small gradients. In addition to simple momentum terms there are more advanced optimizers such as RMSprop (Tieleman and Hinton, 2012) or Adam (Kingma and Ba, 2014), which show faster learning close to these saddle points in high dimensional spaces.

A large set of developed machine learning tools is available to optimize computational goals in neural networks (Bishop, 2006; Bengio, Courville, and Vincent, 2013). This can be used for basic research in cognitive science, by posing a hypothesis about a computational principle, defining a goal function, and optimizing it using gradient descent (Einhäuser and König, 2010). The next section describes how such goal functions can be used to construct models of cortical visual hierarchies.

Unsupervised Learning of Visual Hierarchies

The visual system is hierarchically organized in functional streams with forward, backward, and lateral connections (Felleman and Van Essen, 1991). The development of this modular organization of the visual hierarchy is certainly guided by genetic influences. At the same time, it is unlikely that all detailed visual features, such as specific receptive fields of individual neurons, are only hard wired by genetic expressions without using the activation statistics of nearby neurons. Evidence for this was provided by Hirsch and Spinelli, who exposed cats to controlled horizontal and vertical lines during their development and showed highly selective modification of neuronal connectivity by these environmental conditions (Hirsch and Spinelli, 1971, also see Stryker et al., 1978). Based on this finding, Rauschecker and Singer further analyzed the distributions of neuronal receptive fields for different environmental conditions (Rauschecker and Singer, 1981). They argue that the mechanism of adaptive synaptic plasticity that was postulated by Hebb (Hebb, 1949) could account for the development of these neuronal response properties. Here, I will introduce computational models of representation learning that can explain the development of receptive fields resembling the ones found in the visual cortex.

It is known from physiological experiments that the receptive fields of simple cells in the primary visual cortex are location and orientation selective edge detectors (Hubel and Wiesel, 1962; Marr and Hildreth, 1980). Interestingly, Olshausen and Field showed that the optimization of sparseness of activations while showing natural images to an artificial neural network produces very similar receptive field characteristics (Olshausen, 1996; Olshausen and Field, 1997) under the assumption that an image can be represented as a linear superposition of these receptive field basis functions. This same idea of sparse coding was also theoretically linked to unsupervised learning by information maximization and independent component analysis (ICA, see Bell and Sejnowski, 1997). Based on these milestones, the question is whether there are other computational principles that might also play a role in the cortical microcircuit, which could be optimized by gradient descent learning.

As it turns out, a similar learning rule or computational principle can be found regarding the topographic organization of the visual cortex. In fact, a more detailed analysis of the primary visual cortex reveals a topographic organization of location, orientation, and color selectivity (Inouye, 1909; Hubel and Wiesel, 1962). Furthermore, neurons can be categorized into simple (specific on and off regions in the receptive field, see Movshon, Thompson, and Tolhurst, 1978a) and complex cells (spatially invariant edge detectors, see Movshon, Thompson, and Tolhurst, 1978b). The receptive fields are topographically organized such that cortical columns with a similar receptive field location and similar orientation selectivity are close to each other (Swindale, 1982). Interestingly, Hyvärinen and Hoyer showed that the same type of organization of receptive fields can be achieved in artificial neural networks by the optimization of the sparseness goal in a localized region in a 2-dimensionally spanned feature dimension (Hyvärinen and Hoyer, 2001). Specifically, they maximized the sparseness of the output of complex cells that were hard-wired to receive the squared inputs of locally neighboring simple cells. Figure 1.5A shows the connection weights of the receptive fields of simple cells that were trained with topographic sparseness optimization.

Regarding the implementation of such models, the training with visual inputs is computationally very demanding, especially if several of such layers (or cortical modules) are stacked resulting in a hierarchical architecture. For each position of the visual input there needs to be a large number of neurons that

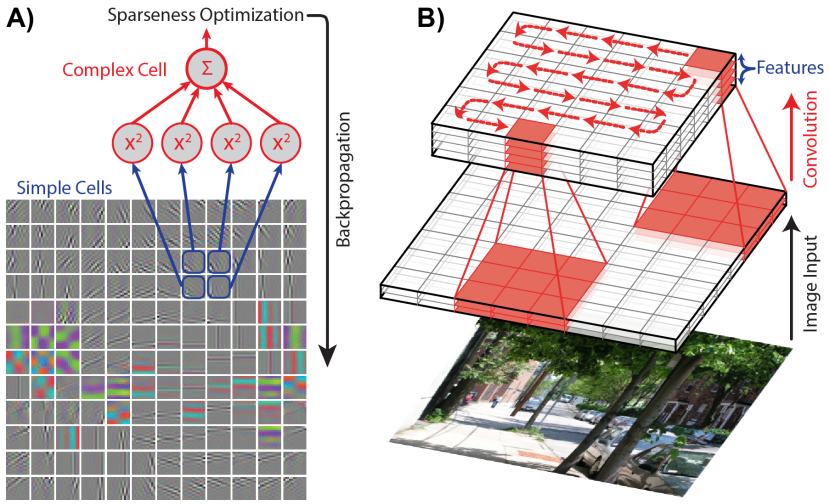


FIGURE 1.5: **Topographic sparseness optimization and convolutional nets: A:** Receptive fields of simple cells are learned by optimizing the sparseness of complex cell activations. **B:** The convolution applies the same weight filter at all positions in the image.

extracts visual features from this location in the image. To alleviate this problem, it is possible to make use of the spatial invariance property of the visual modality by using convolutions across the two dimensions of the images (LeCun and Bengio, 1995). Essentially, this is the same as if many neurons, each with receptive fields at different positions in the visual field, have tied (shared) connection weights, as it is shown in Figure 1.5B. This computational trick can drastically speed up the training process because all parts of the images are used to learn the neuronal weights, although the connectivity is only local.

Autoencoders

Based on the remarkable findings, that the optimization of sparseness, leads to similar receptive fields as found in the visual cortex of the brain, the question remains what other computational principles are implemented by the canonical microcircuits. A more general approach for the description of the above mentioned sparse coding schemes are autoencoder networks that learn not only

forward but also backward connections and minimize the reconstruction error between the reconstructed top-down signal and the original bottom-up input signal (Baldi and Hornik, 1989; Hinton and Zemel, 1994; Bengio and LeCun, 2007; Bengio et al., 2007). The architecture of such a network is shown in Figure 1.6A, where the backward connections are flipped to the top. Noteworthy, the autoencoder learns to compress the input patterns if the hidden layer representation has a lower number of units in comparison to the input layer (Hinton and Salakhutdinov, 2006). In fact, these autoencoders have been shown to be useful for the pretraining of very deep artificial neural networks in machine learning applications (Erhan et al., 2010). In particular, this unsupervised pretraining of deep architectures is achieved by stacking a new autoencoder layer on top of the hidden layer representation of the previously trained autoencoder (Bengio et al., 2007; Bengio, 2009).

A problem that is often faced in machine learning applications is overfitting to the training data, so that the final model does not generalize towards new data (i.e. validation or test data, for review see Tetko, Livingstone, and Luik, 1995; Zhang, 2000). In fact, in unsupervised representation learning a similar problem exists, although the natural sensory inputs have no label and no prediction is performed. For example, an autoencoder that has more units in the hidden layer in comparison to the visible layer would easily learn the identity mapping to perfectly reconstruct the given input. In the temporal coding study, I use such an overcomplete autoencoder to generate a rich feature representation of the sensory inputs, which is necessary for the simulation of binding by synchrony. There are many extensions to the way how autoencoders are trained and how other optimization functions might be added to help prevent overfitting (Vincent et al., 2010). Usually some type of weight decay term is added to the optimization function. This reduces the model capacity, by forcing unused weights to decay towards zero. Alternatively, one can add noise to the input images during training. Based on these noisy inputs, a "denoising autoencoder" would then learn to reconstruct the original image (Vincent et al., 2008). Figure 1.6B shows how such a denoising autoencoder learns to reconstruct the statistics of the underlying original manifold, from which the training examples were drawn, by adding noise to the network inputs x and optimizing the network to reconstruct the original samples. Interestingly, this idea can be further extended to the concept of "dropout", which prevents the coadaptation of the learned representations of neurons by randomly switching

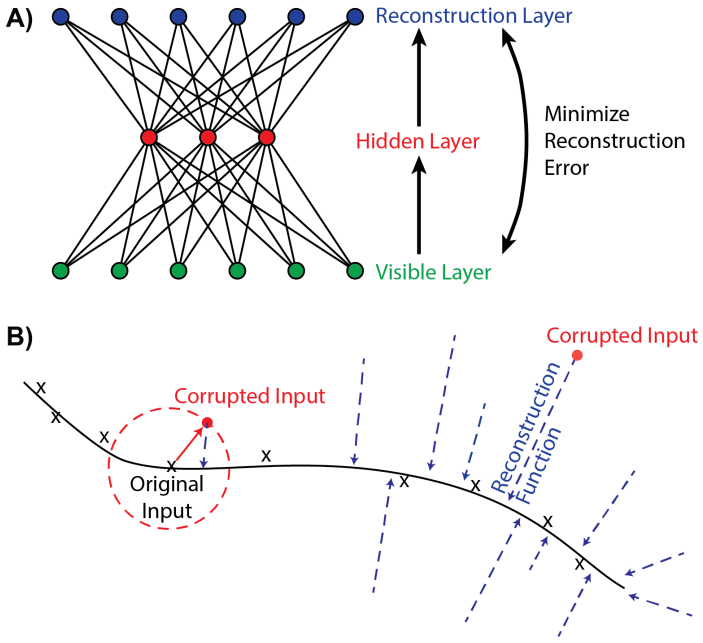


FIGURE 1.6: **Autoencoder and denoising autoencoder. A:** The architecture consists of a visible, a hidden, and a reconstruction layer. The reconstruction error is minimized. **B:** Visualization of the training manifold from which training examples x are drawn. The addition of noise corrupts these original inputs. The denoising autoencoder learns a reconstruction which maps back into the original manifold.

units off during training (Srivastava et al., 2014). In summary, all these methods help to learn overcomplete representations, as the ones in the temporal coding study (chapter 3).

Perceptual Timescales

On behavioral timescales temporal information is very important because relevant objects in the surrounding environment of an individual are usually present over a prolonged time frame. The statistics in this temporal information is used in slow-feature-analysis (Wiskott and Sejnowski, 2002). Even more interesting

is the application of this idea to the ventral visual stream (Goodale and Milner, 1992). For example, Wyss and colleagues showed that a temporal stability objective can produce place fields in higher layers of a hierarchical network (Wyss, König, and Verschure, 2006) if the visual inputs are generated from a randomly moving robotic agent.

As illustrated in the beginning of this general introduction in Figure 1.1, several timescales are involved when the brain processes information from the environment. In this regard, the above mentioned temporal stability objective would be classified at the medium behavioral timescales because it employs the fact that the robotic agent perceives the environment in a temporally stable way. The unsupervised learning process of these place cells would result in changes of the network connectivity on slow timescales (i.e. gradual changes by small adaptation of weights for each received image from the environment). The motivation for the temporal coding study (chapter 3) is to further extend such models also with dynamic network organization on fast timescales. In this regard, I use the Kuramoto model of coupled phase oscillators in combination with such rate based models (Finger and König, 2013), as it was described in section 1.1. The activation of these oscillators is determined by rate-based feed-forward connections from natural visual stimuli. As a result, the model covers slow (learning of structure), medium (behavioral perception and network activation), and fast timescales (oscillatory network interactions).

To summarize this section, deep learning of network hierarchies has been very successful in machine learning tasks. Interestingly, the unsupervised pretraining of such hierarchies by stacking autoencoders is related to the unsupervised learning of sensory representations in the brain. But many questions remain regarding the computational principles that are at play in biological neural networks and how these could be related to artificial neural networks. In fact, the machine learning research is focused almost exclusively on these rate-based neuron models. In the next section, I will present the results of the temporal coding study (Finger and König, 2013), which extends such rate-based neurons with phase variables to model dynamic coordination in these types of networks. This could further improve deep learning architectures, as introduced in this section, with the ability to self-organize the effective functional connectivity across multiple network nodes.

1.4 Temporal Coding and Binding in Distributed Representations

The first section of this general introduction presented the overall modeling approaches that are used in this thesis. Next, the second section showed how the simulation of dynamic functional interactions can be applied to the connectome. The third section introduced mechanisms how this structural connectivity can be learned by applying gradient descent optimization to the functionally activated network. This section now combines these approaches to analyze the computational principle underlying functional dynamics on fast timescales and how this helps dynamic coordination and binding of distributed representations.

In the previous section we saw that unsupervised learning is usually applied to rate-based neural networks. As long as these hierarchical neural networks are just performing feature extraction with a fixed task it is usually sufficient to have a feed-forward architecture of several layers of neurons. However, as soon as the flexibility of tasks and the instant adaptation to a new behavioral situation is required, we can assume that a simple feed-forward architecture is not sufficient. Instead a highly complex dynamic system with interacting modules with recurrent connections is necessary to achieve a high level of flexibility. This raises the question, how these local processing circuits are coordinated in a complex network like the human brain so that they each fulfill their corresponding role, leading to the emergence of task relevant computations (Malsburg, Phillips, and Singer, 2010). In this context, it has been suggested that fast oscillations in the frequency bands starting from the delta range (1-4 Hz) up to the gamma range (30-70 Hz) could provide the necessary mechanisms to dynamically coordinate spatially distributed activations in the brain (Gray et al., 1990; Fries, 2015; Herrmann et al., 2015; Maris, Fries, and Ede, 2016; Michalareas et al., 2016).

One such organizational principle is concerned with the binding problem (Malsburg, 1994; Malsburg, 1995). The question of coordination is very relevant in these distributed representations on the level of local cell populations but also on the level of brain areas. Figure 1.7 shows an illustration of two stimuli that would each activate a specific population of neurons. If the representations of both stimuli have to be activated at the same time (i.e. due to holding both

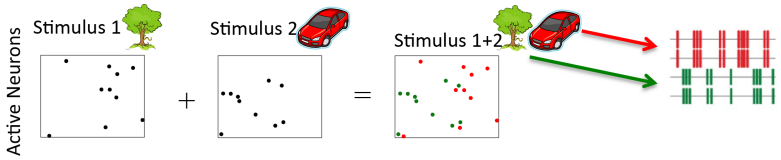


FIGURE 1.7: **Binding by synchrony.** If the neuronal representation of two stimuli (green tree and red car) are activated at the same time the association of each neuron could be coded with the timing of spikes or the phase of their oscillation.

items simultaneously in working memory, or due to perceiving both stimuli simultaneously), a rate-coding of the neurons activation could not distinguish between the stimuli locally in the network. In contrast, the association of neurons belonging to the same object and dissociation between neurons of different objects could be represented with the timing of their action potentials.

Theories of dynamic coordination in the brain

Before diving deeper into models of temporal coding and binding by synchrony, I will first relate the idea of dynamic coordination to other, more general, normative models of neural information processing. In contrast to unsupervised learning as described in section 1.3, these more general frameworks also model contextual interactions or internal states and thereby go beyond a simple feed-forward architecture of sensory perception.

Neural processing systems can be described in a normative way using information theoretic methods. For example, Linsker described perceptual networks in terms of information maximization (Linsker, 1988). Specifically, this "infomax" algorithm describes neurons as maximizing the mutual information between the receptive field input and the neuronal output activation. Later, this idea was extended to a more general framework, called "coherent infomax", which decomposes the information content conveyed by a neural processing unit into four components (Kay and Phillips, 1997; Kay, Floreano, and Phillips, 1998; Kay and Phillips, 2011). Figure 1.8 visualizes these four information theoretic components in terms of mutual information between the units activation, the receptive field input, and the contextual input. The

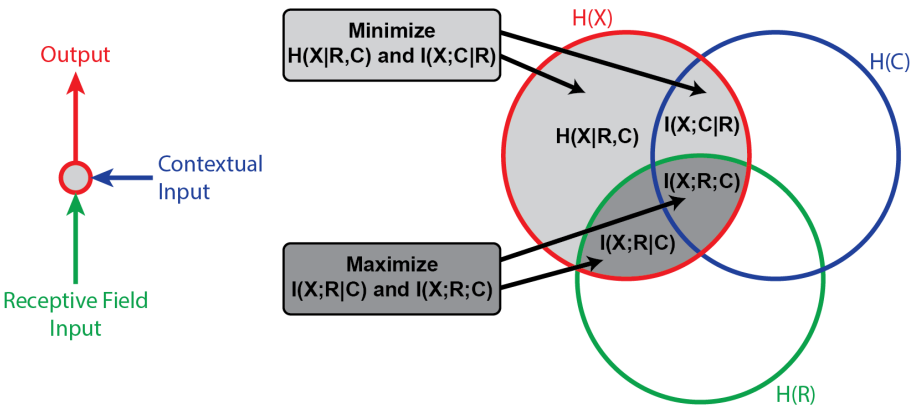


FIGURE 1.8: **Coherent infomax.** A local processing unit (red) is visualized with receptive field inputs (green) and contextual field inputs (blue). Coherent infomax decomposes the information entropy $H(X)$ that is conveyed through the output of the unit. Different weightings can be used for optimizing the four information components in the output: the conditional Shannon entropy $H(X|R,C)$, the conditional mutual information $I(X;C|R)$ and $I(X;R|C)$, and the three-way mutual information $I(X;R;C)$. Figure based on Kay and Phillips, 1997.

units can then be optimized while applying different weightings to the optimization of these four information theoretic terms. Usually, the two terms that represent the receptive field input that is not matching with the contextual inputs ($I(X;R|C)$) and the overlap between the receptive field input and the contextual inputs ($I(X;R;C)$) are maximized, while the other two terms ($H(X|R,C)$ and $I(X;C|R)$) are minimized. Importantly, the contextual input allows the units to adapt to the current context of computation in the network by amplifying or attenuating the effect of the receptive field input onto the output of the unit. Therefore, the theory of coherent infomax describes a form of dynamic coordination in the brain as a contextual modulatory effect on the transmission of information.

Another general theory of neuronal computation is the free-energy-principle proposed by Friston (Friston, 2010). Although a detailed description of the free-energy-principle would go beyond the scope of this thesis, it should be

noted that it describes the computational principle of neural processes as minimizing a free-energy term. This free-energy can be framed in terms of predictive-coding by describing top-down connections as transmitting predictions to lower layers, while bottom-up connections transmit the corresponding prediction errors. Therefore, in contrast to coherent infomax, the free-energy principle also includes loop back connections, which could even be extended to the complete sensory-motor loop.

Both coherent infomax (Kay and Phillips, 1997) and the free-energy principle (Friston, 2010) provide a very high level description of computational goals that the brain could be optimizing. One might ask, what knowledge besides these general theories is still missing to fully understand the computational mechanisms of our brain. One aspect that comes to mind, and that is left unexplained by these theories, is the required amount of structural connectivity between the local processors in the network that is needed to execute high level cognitive tasks. For example, the free energy principle describes the internal states of a unit as dependent on the previous internal states and the sensory states (or inputs from lower layers) without specifically mentioning how the sensory input would be grouped into local processing modules to reduce the amount of required wiring. One might ask, which network topology could potentially support the needed rich flexibility such that all pairwise network interactions could be made possible if needed. This leads to the second aspect that might still be missing in these general theories. That is the role of neural oscillations. Both theories do not explicitly assume brain oscillations as mechanisms of dynamic brain organization, because both theories are formulated in information theoretic or probabilistic terms. However, brain oscillations could explain how the complexity and flexibility can be implemented by our brain, without requiring an intractable amount of connections between all pairs of processors that might need to exchange information (Singer and Gray, 1995). Specifically, oscillations might be very relevant in routing information through brain regions that act as hubs in the network (Womelsdorf et al., 2007; Akam and Kullmann, 2010). In fact, dynamic switches in effective brain connectivity at these network hubs could allow a vast expansion of cognitive capabilities (Malsburg, Phillips, and Singer, 2010). In the following section, I will introduce the relevant neurophysiological background of these oscillatory processes in the brain and how they may subserve computational principles.

Dynamic Organization by Temporal Coding

Temporal coding and synchrony of oscillations play a major role in the coordination of activations between individual cells, but also across cortical columns or brain areas. On the microscopic scale, this is illustrated by the influence of spike-timings on neural plasticity mechanisms (Markram et al., 1997) and by evidence that global stimulus properties are reflected in the local inter-columnar synchronization (Gray et al., 1989). On larger spatial scales, the importance of temporal coding is indicated by the important role of cortical oscillations for top-down processing and attention mechanisms (Engel, Fries, and Singer, 2001; Fries et al., 2001; Jensen, Kaiser, and Lachaux, 2007; Lakatos et al., 2008), as well as for memory functions (Lisman and Idiart, 1995; Klimesch, 1997; Klimesch, 1999).

Of specific interest in this thesis is the role of oscillations in relation to the binding problem (Malsburg, 1994; Malsburg, 1995). Von der Malsburg hypothesized that Gestalt principles of perception (Smith, 1988) might be grounded in neurophysiological mechanisms that bind distributed neural representations by their synchronization (Malsburg, 1981; Malsburg and Schneider, 1986). Evidence for this hypothesis was provided by Gray and colleagues, who found that global stimulus properties were reflected in the local inter-columnar synchronization in the gamma (40-60 Hz) frequency range (Gray et al., 1989, also see Engel et al., 1990). Engel and colleagues provided further evidence for scene segmentation by temporal coding (Engel, König, and Singer, 1991) and for interhemispheric synchronization, indicating that such mechanisms also play a role across cortical areas (Engel et al., 1991). Overall, it was speculated that coherence might play a functional role for binding of the activations in distributed neural representations (Malsburg and Schneider, 1986; Singer, 1999).

This dynamic organization of larger brain areas and cortical columns might be grounded in the temporal coding at the individual neurons by precise spike-timings. There is evidence that information processing in the brain relies at least in some cases on the precise timings of action potentials (Engel et al., 1992; Victor and Purpura, 1996), although the specific contribution of firing rate coding in comparison to temporal coding is hard to quantify. In this respect, Mainen and colleagues showed that timing of action potentials reveals information about a perceived stimulus (Mainen and Sejnowski, 1995). More

recently, Gutkin and colleagues provided evidence that constant stimulation leads to imprecise timings of action potentials, but for temporally changing stimulation the neurons act as threshold elements leading to precise spike timings (Gutkin, Ermentrout, and Rudolph, 2003). In consequence, we might suppose that intrinsic oscillatory processes enable temporal coding with spikes throughout the brain, because intrinsic oscillations automatically result in constantly changing outputs and inputs in all brain areas and not only in low level sensory regions.

The importance of stimulus-locked spike trains was shown for binaural sound localization, which is performed by estimating the time difference between the arrival of sound waves between the left and right ear (Carr and Konishi, 1988). A time-delay neuronal network, called the Jeffress model, was proposed for estimating this interaural time difference (Jeffress, 1948). In this network, neurons act as coincidence detectors between two reciprocal delay lines from the left and the right ear. This model of sound source localization was later experimentally validated in the barn owl (Carr and Konishi, 1988). In this regard, I implemented a spike-based model of sound source localization with a neuromorphic cochlea (Finger and Liu, 2011). However, this was work preceding my doctorate studies and is therefore not part of this thesis. Further evidence for the importance of the precise temporal alignment of action potentials are neurophysiologically realistic models of synaptic plasticity. Here, experiments revealed an asymmetric learning rule that strengthens synaptic weights in the case of a preceding presynaptic spike before a postsynaptic spike, and conversely a weakening if the temporal order is reversed (Markram et al., 1997). Importantly, it was shown that this spike-timing dependent plasticity (STDP) creates competition between different synaptic inputs (Song, Miller, and Abbott, 2000). In particular, all input synapses of a neuron compete for the control of the timing of the postsynaptic action potentials.

Regarding the memory consolidation study (chapter 4.2) that will be introduced in more detail in section 1.5, the theta rhythm in the hippocampus plays an important role for memory functions. Specifically, O'Keefe and Nadel found evidence for temporal coding by spike-timings and cortical oscillations in the theta rhythm in the hippocampal network and it's communication with the neocortex (O'keefe and Nadel, 1978). Later, O'Keefe and Recce found that activity of place cells advances to earlier phases of the theta cycle when a rat walks through the corresponding place field (O'Keefe and Recce, 1993).

This so called theta-phase-precession was later analyzed in more detail revealing an important role for the compression of temporal sequences (Skaggs and McNaughton, 1996b). Based on these findings, theories about the coding of working memory items in the hippocampus within these theta phases have been developed (Buzsáki, 2002; Jensen and Lisman, 2005; Lisman, 2005). For example, Lisman speculates that cross-frequency coupling between the theta and the gamma rhythms might implement a working memory buffer (Lisman, 2005). Further indication for this theory is provided by Alvarez, who showed that the visual short-term memory has a capacity limit of approximately four or five objects and that this limit is filled not only by the number of visual objects, but also by the number of task relevant object properties (Alvarez and Cavanagh, 2004). This memory buffer could allow to reactivate several memory items within one theta cycle, so that long-term potentiation (LTP) between the representations of these items could take place (for a review see Squire and Wixted, 2011). Overall, it was proposed that the theta rhythm might play a crucial role in the coordination of communication between the hippocampus and the neocortex (Buzsáki, 1996; Sutherland and McNaughton, 2000; Sirota et al., 2008; Benchenane et al., 2010).

In summary, oscillatory processes play a major role in communication processes in the brain. This conjecture is further supported by the fact that activated neuronal populations have the intrinsic property to oscillate (Silva, Amitai, and Connors, 1991; Hutcheon and Yarom, 2000). In general, Fries hypothesized that the communication between brain areas is flexibly modulated by coherence (Fries, 2005; Fries, 2015). It seems that synchrony or temporal coherence is responsible for the coordination of brain activations and this role might be grounded in neurophysiological properties of intrinsic oscillations of neural populations and learning rules that depend on precise spike-timings.

Unsupervised Binding in Distributed Representations

As described above, the oscillatory nature of brain signals facilitates important functional roles, such as coordination, attention, memory, and binding of distributed representations. But in artificial neural networks information is usually represented with only rate-based units. These units code the activation only as a scalar value and do not integrate over an oscillatory limit-cycle. For

machine learning applications this is reasonable in order to reduce the computation time that would otherwise be wasted in the integration of differential equations. However, this is a drastic simplification of real biological networks. A legitimate concern is that this simplification loses several key mechanisms, which might be functionally relevant and overall beneficial for the computation in the network. One such mechanism, which is lost, is the ability to dynamically bind different activated features of an object in the distributed representation in the network. Therefore, here I will introduce an implementation that integrates this mechanism into simple rate-based neuron models.

As already illustrated in Figure 1.7, binding is important for visual perception. In fact, binding by synchrony has been proposed as a mechanism to accomplish this within distributed neural representations (Milner, 1974; Gray and Singer, 1989; Engel, König, and Singer, 1991; Gray, 1999). Based on the findings of coherent oscillatory activity in the cat visual cortex (Gray and Singer, 1989; Engel et al., 1990), Sompolinsky and colleagues showed that a neural network of phase oscillators is capable of processing global stimulus attributes with binding by synchrony (Sompolinsky, Golomb, and Kleinfeld, 1990). More recent work on binding by coupled oscillators in neural networks is based on the "local excitatory global inhibitory oscillator network" (LEGION) or similar variants of this model (Wang and Terman, 1997; Chen and Wang, 2002). In both the Kuramoto based models (Sompolinsky, Golomb, and Kleinfeld, 1990) and the LEGION models (Wang and Terman, 1997) the connectivity that couples the oscillator nodes is determined explicitly by hand designed features.

This motivates the question whether a similar model of binding by synchrony and image segmentation is feasible in the setting of unsupervised representation learning from sensory statistics as introduced in section 1.3. The temporal coding study (chapter 3) investigates such a combination with a model of coupled oscillators within a network of unsupervised learning from natural image statistics (Finger and König, 2013). How this model combines activation and phase variables was already described in section 1.1 (equation 1.2). In this study the neural activation (rate) is calculated with convolutional forward connections that were trained by an autoencoder objective function as described in section 1.3. Furthermore, the model makes use of sparse horizontal connections that are based on the activation statistics of the corresponding network nodes. These horizontal connections are coupling Kuramoto oscillators, which are activated by the forward connections based on their receptive

fields. Importantly, the horizontal connections are generated with the assumption that correlated units have a high probability to be coupled by synchronizing connections, while anticorrelated units are more probable to be connected by desynchronizing connections (Finger and König, 2013).

The evaluation of the simulated phase variables utilizes hand labeled image segments that correspond to high level objects as ground truth. As a baseline we also evaluate the same segmentation masks on simulations based on other images to account for the general synchronization behavior of neighboring neurons. Finally, the results show that this normative model is able to dynamically bind the neuronal populations that encode properties of the same visual object.

Overall, the study demonstrates how the extension of rate-based networks with phase variables bears the potential for dynamic organization of the activated neural populations. These findings highlight the importance of temporal coding by oscillatory interactions in the brain. Moreover, the results validate that the approach of a second variable to code contextual relationships in deep networks might be feasible also for improvements of machine learning algorithms. However, for the application of such models more work needs to be done to reduce the required computational time to simulate the phase interactions. In this regard, autoencoders with complex valued units could be used to replace the Kuramoto model (see general discussion 5.1). Further extensions of this model are also possible regarding more layers and phase variables that could synchronize different modules (see general discussion 5.3).

1.5 Relation to Behavior and Cognition

As discussed in the above sections, the brain function is highly dependent on the interactions between structural, and functional connectivity, and on the dynamic coordination of this functional activation. These interactions lead to important behavioral and cognitive consequences. The mechanism of dynamic coordination in the brain might support the high flexibility of cognitive functions. This is manifested in the ability to switch very fast between executing different tasks by utilizing the appropriate brain processes. An example of this utilization of multiple brain areas and their interaction is multimodal integration of sensory information (e.g. integration of visual and auditory information). Another important cognitive function, which requires the coordination between multiple brain areas is the consolidation of newly acquired knowledge and experiences into long term memory. Therefore, this section will relate the topic of structural connectivity, functional activation, and dynamic interactions to multimodal integration and memory consolidation.

Sensory Integration or Alternation

Perceived sensory information needs to be integrated for a correct interpretation of the environment. A prerequisite for this integration is a learned compressed representation of the raw sensory inputs (compare section 1.3). For integration of information from different modalities, a mapping between these uni-modal representations needs to be learned. Importantly, the uni-modal perception of a physical property is usually not perfect, but it is corrupted by sensory noise and hence associated with a corresponding uncertainty of the estimation. If the brain has the ability to estimate these uni-modal uncertainties, then the integration of both modalities could reduce the uncertainty in the combined information. This poses the question, how the brain integrates information from two or more modalities and how this is achieved by the coordination of communication between the corresponding cortical areas.

In this regard, Ernst and Banks analyzed the integration behavior when humans explore object dimensions with the visual and haptic senses (Ernst and Banks, 2002). They showed that humans combine the information from both

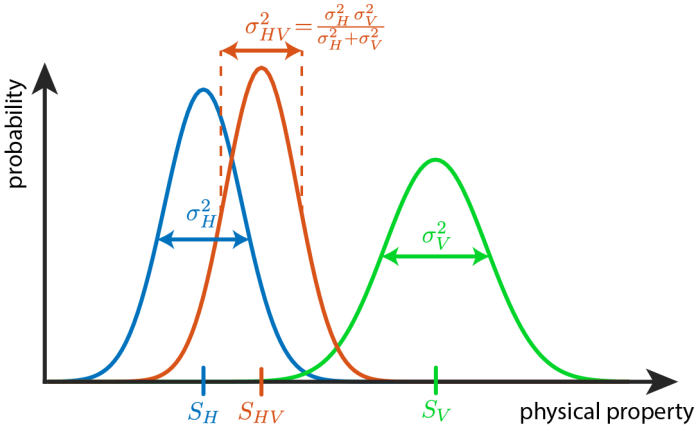


FIGURE 1.9: **Maximum likelihood estimation in sensory integration.** The haptic input is described by the blue Gaussian with mean S_H and variance σ_H. The visual input is described by a green Gaussian with mean S_V and variance σ_V. The maximum likelihood estimate (red) is described by a Gaussian with mean S_{HV} and an overall smaller variance σ_{HV}.

modalities in a way that is similar to a maximum likelihood integrator. Specifically, they assume that the uni-modal estimates are corrupted by independent Gaussian noise with a given standard deviation for each modality as shown in Figure 1.9. Assuming a uniform Bayesian prior, this results in a maximum likelihood estimate of the integrated information that is again described by a Gaussian.

Cue combination from different sensory modalities was also studied by synchronized presentation of auditory and visual stimuli (Körding et al., 2007). Körding and colleagues showed that the temporal synchrony of presentation has an effect on the integration process. This shows that synchrony is not only relevant within the brain, but also that the temporal synchrony of external events in the environment is of high relevance to sensory integration and learning. For example, Pairse and colleagues argued that the brain infers causality from a correlation between different modalities (Parise, Spence, and Ernst, 2012). This idea also relates to the principles of normative learning, where an objective function that maximizes temporal stability of the activations was shown to produce place fields in higher hierarchical layers (Wyss, König, and

Verschure, 2006). One could hypothesize that the same optimization function applied to a layer with multimodal inputs could also learn cue integration depending on synchronous presentation.

More recent research has demonstrated that multisensory integration is not statically wired from birth but learned by simultaneous activation of the sensory modalities. For example, it has been shown that children do not integrate different modalities, but that multisensory integration only develops late in humans (Gori et al., 2008; Burr and Gori, 2012). This learning process was further analyzed by Ernst with a device that can present arbitrary information into the visual field through a horizontal mirror while participants had to perform haptic manipulations below this mirror (Ernst, 2007). This device was used to code information about the stiffness of objects as brightness into the visual sensory inputs. The results are evidence that subjects learned the sensory integration of the new information from the statistical co-occurrence during training.

In the sensory integration study (chapter 4.1) we use such a sensory augmentation device to investigate multisensory integration (Goeke et al., 2016). In relation to the topic of this thesis, it is important to notice, that a sensory integration task, specifically with a sensory augmentation device, requires a high level of dynamic interaction between brain areas, that are usually not interacting. A sensory augmentation device relays information to an existing sensory area, which is then stimulated by completely new patterns that need to be integrated into the existing organization of the connectome. Therefore, these new inputs need to be routed to the relevant brain regions where they can be integrated into the existing pathways from other senses with related information. To investigate these mechanisms, we use a tactile belt and a rotation platform to stimulate simultaneously the tactile and the vestibular sense with controlled inputs. The subjects had to indicate their estimated rotation angle, for which they could either employ the tactile or the vestibular or both senses. This sensory integration task requires constant communication between brain areas because the rotation angle can only be obtained from the vestibular sense by temporal integration of angular acceleration. Our results suggest that participants are flexible in switching between the vestibular and the tactile input. Only some of the participants were able to combine both inputs to further improve the estimated rotation angle.

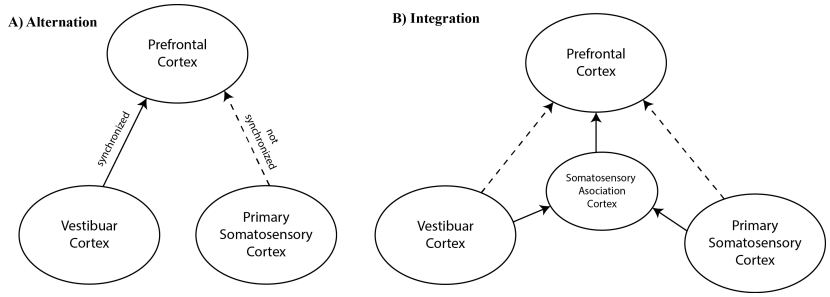


FIGURE 1.10: **Bimodal integration or alternation as multi-site interaction in the brain.** (A) Before learning of bimodal representations in the somatosensory integration areas, the rotation information is used from either the vestibular or the tactile condition depending on synchronization of brain areas. (B) After learning of bimodal representations in the somatosensory association cortex, both modalities can be integrated unconsciously.

We use a Bayesian alternation model to account for the differences in integration performance or preferences for one of the modalities for different subjects. This model can be adjusted to use specifically more information from one of the two modality while neglecting the other (Nardini et al., 2008). Our results show that this Bayesian alternation model better predicts the response of subjects especially if we use the subjective answers of the participants in which they indicate their preference for one of the two modalities.

A possible interpretation of these results with respect to the topic of this thesis is illustrated in Figure 1.10A. When subjects wear the sensory augmentation device for the first time, we can assume that there are no direct structural connections that would directly connect the newly augmented information from the primary somatosensory cortex to the existing rotation information in the vestibular cortex. Therefore, a pathway for the integration of both cues would probably involve multiple network hops, probably through the prefrontal cortex. The information from both senses is probably read out at this higher level in the cortical processing hierarchy where the final judgment for the response decision is made. Depending on the synchronization of this higher level cortical area, with one of the two lower level sensory regions, the response would be based on either one of the two modalities.

We can further speculate that a more direct mapping between both early sensory areas could be learned if the sensory augmentation device would be used throughout an extensive period of time. Such an early integration of the rotation cue from the vestibular and tactile inputs could be learned in a similar way as the normative models that were described in section 1.3. Finally, this would result in multimodal sensory representations (probably in the multisensory association areas) as visualized in Figure 1.10B. Overall, this hypothesis demonstrates the important interplay between dynamic organization of functional brain interactions and the learning of structural connectivity based on sensory statistics.

Memory, Oscillations and Sleep

The previous section introduced dynamic multi-site communication in the context of multi-sensory integration and alternation. Another cognitive mechanism that is also related to multi-site interactions are memory functions, such as consolidation and retrieval of associative memory items (Moscovitch, 1995). Since Scoville and Milner reported about the famous H.M. patient, the hippocampus is believed to play a crucial role in these memory functions (Scoville and Milner, 1957). In this regard, Nadel and colleagues review a "standard model" of memory formation, in which the hippocampus and related networks in the medial temporal lobe act as a temporal storage of new memories, which are only later consolidated into a more persistent storage (Nadel and Moscovitch, 1997). Specifically, the consolidation of newly acquired knowledge is thought to be mediated by the transfer of episodic memories from medial temporal to neocortical networks (Squire and Zola, 1997; Eichenbaum, 2000; Remondes and Schuman, 2004). Therefore, the consolidation and retrieval require some form of coordination of the communication between the involved brain areas.

Episodic events are not immediately transferred into long-term memory. Instead, it is thought that the consolidation takes place mostly due to repetitive activation of the corresponding neural representations (Louie and Wilson, 2001; Skaggs and McNaughton, 1996a). Such activations might also occur during sleep (Born and Wilhelm, 2012), which motivates the memory consolidation study (chapter 4.2) to compare brain activations during retrieval of a

post-sleep with a pre-sleep baseline (Köster et al., 2016). Therefore, in the following, I will review the relations between sleep, memory functions, and cortical oscillations.

There is evidence that EEG oscillations in the alpha and theta band reflect memory performance (Klimesch, 1999). In this regard, alpha activity might serve the role to suppress activations of brain regions that are not necessary for the given task (Foxe and Snyder, 2011; Klimesch, 2012). Therefore, alpha activity could be interpreted as fulfilling a role in global gating of information in the brain so that important information reaches the relevant brain areas (Jensen and Mazaheri, 2010). Moreover, theta activity was shown to be important in the hippocampal system and might play a crucial role in memory functions by coordinating the communication with neocortical brain areas (Sirota et al., 2008; Benchenane et al., 2010).

In regard to the microscale network function and the oscillatory activity in the hippocampus, most evidence was obtained using single-unit recordings in animal experiments (O'Keefe and Dostrovsky, 1971; O'keefe and Nadel, 1978; McNaughton, Barnes, and O'keefe, 1983; Foster and Wilson, 2006). These studies revealed that the hippocampus plays not only an important role in memory formation but also in navigation. More precisely, O'Keefe and Dostrovsky discovered that single units in a rat's hippocampus were active specifically when it passed a specific location in a maze (O'Keefe and Dostrovsky, 1971), and therefore named these units place cells. Interestingly, this indicates that the two seemingly different functions of navigation and memory formation are linked. Based on this, Buzaki and Moser hypothesized that the hippocampus evolved from a navigation to a more general episodic memory system that allows to navigate in mental space in a similar way to the navigation in physical space (Buzsáki and Moser, 2013). In fact, this relation of the human memory system to place cells, motivates further interesting hypotheses about temporal coding and dynamic organization that will be presented in the following paragraph.

As already described in section 1.4, it has been suggested that theta oscillations in the hippocampal system play the role of a memory buffer (Buzsáki, 2002; Jensen and Lisman, 2005; Lisman, 2005). This memory buffer can rapidly store and recall newly acquired memory items in a sequential order (Treves and Rolls, 1994). Therefore, the hippocampus acts as a short-term memory and is

also highly involved in the consolidation of new memory items into long-term storage (Buzsáki, 1989). There is growing evidence that this consolidation of new memory items into long term memory is mediated by active replay, meaning the reactivation of stored memory sequences (Euston, Tatsuno, and McNaughton, 2007; Peyrache et al., 2009). This replay could be elicited by active repetition in the awake state, but also importantly by automatic replay during sleep. In this respect, Born and Wilhelm reviewed the evidence for the consolidation of newly acquired associative memories during sleep (Born and Wilhelm, 2012). They hypothesize that this consolidation takes place during slow-wave sleep where memories are reactivated to be consolidated. Furthermore, Gais and colleagues found that the activation during retrieval of memory items is shifted from medial temporal to neocortical networks after a period of sleep (Gais et al., 2007; Takashima et al., 2006).

In addition to this "active system consolidation hypothesis" (active replay facilitating consolidation) described above, it has also been suggested that sleep downscales overall synaptic strengths back to a baseline level. Tononi and Cirelli reviewed the evidence for this synaptic homeostasis hypothesis (Tononi and Cirelli, 2006). They state that this downscaling of synaptic weights downregulates average firing rates leading to more efficient energy use. Furthermore, they argue that it has benefits for learning by freeing unused network capacities for new memories and leads to a more efficient use of the structural connectivity. Noteworthy, this hypothesis of synaptic downscaling is not mutually exclusive with the hypothesis of consolidation by replay and both effects could be facilitated by sleep (Ellenbogen, Payne, and Stickgold, 2006; Tononi and Cirelli, 2006; Diekelmann and Born, 2010).

To gain further insights into the role of cortical oscillations during memory consolidation and retrieval in relation to sleep, we recorded the EEG of participants during an associative memory task in the memory consolidation study (see chapter 4.2). We specifically tested associative memory retrieval before and after a period of sleep to measure differences that arise due to the consolidation during sleep (Köster et al., 2016). As a baseline we used a control group that was awake between the two retrieval sessions. The experiments demonstrated a decrease in theta power during the retrieval task in the sleep group in comparison to the wake group (when contrasting the post with the pre retrieval session). Overall, these results are in line with the hypothesis that after sleep memories were consolidated into neocortical networks so that the retrieval is

less dependent on theta modulated communication with the hippocampus. For a more detailed discussion of these results and further analyses of the alpha and gamma frequency ranges see the manuscript in chapter 4.2.

There are three overarching ideas that relate this study to the topic of this thesis: First, the study illustrates the concept of dynamic organization of functional interactions as the alpha frequency could be gating the information flow by suppression of brain areas that are unrelated to the task (Jensen and Mazaheri, 2010).

Second, theta oscillations are related to the communication between the hippocampal and neocortical networks, demonstrating how oscillations in specific frequencies might modulate effective communication between brain areas. Moreover, these theta oscillations contain multiple gamma cycles each coding a different item in a working memory buffer (Lisman, 2005). This is related to the temporal coding study, in which a computational model with phase variables simulates binding by synchrony. Specifically, I demonstrate in the temporal coding study (chapter 3) that neurons encoding the same visual object were synchronizing to the same relative phase angle. My computational model of binding by synchrony could now be interpreted in the light of this working memory buffer with phase variables in the theta frequency representing items in a working memory buffer.

A third relation between this memory consolidation study and other topics of this thesis is in regard to the unsupervised learning from sensory statistics. The associative memory items that were used in this study were visually presented objects together with a specific background color. This memory consolidation task could be implemented in a neural network with unsupervised learning using an autoencoder objective. In this regard, the relation between neurophysiologically realistic networks with spike-timing dependent plasticity and autoencoder learning rules will be discussed further in chapter 5.1.

Functional Role for Behavior and Cognition

The experimental results of the two behavioral studies (the integration task and the associative memory task) demonstrate that cortical oscillations play an important role in cognitive functions. The sensory integration task demonstrates

the flexibility in brain communication between multiple brain areas. The results show that subjects are able to route the augmented sensory information about their rotation to the relevant brain regions to make an informed decision. Interestingly, the Bayesian alternation model can explain the behavioral results of the subjects better than a model of optimal Bayesian integration. This indicates that subjects can use either one of the two modalities by switching the communication channels between relevant brain areas.

The memory consolidation study (chapter 4.2) highlights the important role of cortical oscillations for the consolidation of newly acquired knowledge. We showed that EEG oscillations during retrieval can index changes that took place in the structural connectivity by consolidation (learning processes). Therefore, our study demonstrates how flexibility in cognitive functions is facilitated by functional interactions on fast oscillatory timescales while the structural connectivity on this full brain level remains almost static.

1.6 About this dissertation

In the previous sections, I have presented the background and motivation of the four peer-reviewed published articles, which are attached in the next chapters. Specifically, it was shown that dynamic coordination of the distributed computations performed by our brain is important on the microscale between single cells, but also on the macroscale between brain areas. This coordination is fostered by the interplay between brain structure and brain function that can be studied at different levels of abstraction.

This interplay between brain structure, activation and dynamics on different levels of abstraction is investigated in the four different manuscripts. In chapter 2, starting from the intermediate level that is concerned with coordination of neural activity between different brain areas, I use a computational model of functional brain connectivity that is informed by empirical DTI. In chapter 3 the focus is on the microscopic level with models of normative learning extended with temporal coding and how sensory information is represented in the neural network. Finally, chapter 4 presents two manuscripts that relate these ideas to behavioral and cognitive functions: One study in the context of multisensory integration, and another study analyzes oscillatory brain activity in relation to associative memory. All these three chapters conclude with a summary section that relates the corresponding findings to the overall topic of this dissertation.

In summary, the computational models and empirical experiments performed in this thesis improve our understanding of the computational principles that allows the brain to flexibly execute a large diversity of tasks. Therefore, the final general discussion (chapter 5) will highlight some of these overarching consequences that result from the published manuscripts. The discussion chapter will also present some new ideas, which were developed during the work on these interesting problems but did not fit into the frame of a manuscript yet.

Chapter 2

Functional Interactions in the Structural Connectome

2.1 Modeling of Large-Scale Functional Brain Networks Based on Structural Connectivity from DTI

This section was published as a peer-reviewed article: Holger Finger, Marlene Bönstrup, Bastian Cheng, Arnaud Messé, Claus Hilgetag, Götz Thomalla, Christian Gerloff, and Peter König (2016). "Modeling of Large-Scale Functional Brain Networks Based on Structural Connectivity from DTI: Comparison with EEG Derived Phase Coupling Networks and Evaluation of Alternative Methods along the Modeling Path". In: *PLoS Comput Biol* 12.8, pp. 1–28. DOI: 10.1371/journal.pcbi.1005025. URL: http://dx.doi.org/10.1371%2Fjournal.pcbi.1005025

Abstract

In this study, we investigate if phase-locking of fast oscillatory activity relies on the anatomical skeleton and if simple computational models informed by structural connectivity can help further to explain missing links in the structure-function relationship. We use diffusion tensor imaging data and alpha band-limited EEG signal recorded in a group of healthy individuals. Our results show that about 23.4% of the variance in empirical networks of resting-state

functional connectivity is explained by the underlying white matter architecture. Simulating functional connectivity using a simple computational model based on the structural connectivity can increase the match to 45.4%. In a second step, we use our modeling framework to explore several technical alternatives along the modeling path. First, we find that an augmentation of homotopic connections in the structural connectivity matrix improves the link to functional connectivity while a correction for fiber distance slightly decreases the performance of the model. Second, a more complex computational model based on Kuramoto oscillators leads to a slight improvement of the model fit. Third, we show that the comparison of modeled and empirical functional connectivity at source level is much more specific for the underlying structural connectivity. However, different source reconstruction algorithms gave comparable results. Of note, as the fourth finding, the model fit was much better if zero-phase lag components were preserved in the empirical functional connectome, indicating a considerable amount of functionally relevant synchrony taking place with near zero or zero-phase lag. The combination of the best performing alternatives at each stage in the pipeline results in a model that explains 54.4% of the variance in the empirical EEG functional connectivity. Our study shows that large-scale brain circuits of fast neural network synchrony strongly rely upon the structural connectome and simple computational models of neural activity can explain missing links in the structure-function relationship.

Author Summary

Brain imaging techniques are broadly divided into the two categories of structural and functional imaging. Structural imaging provides information about the static physical connectivity within the brain, while functional imaging provides data about the dynamic ongoing activation of brain areas. Computational models allow to bridge the gap between these two modalities and allow to gain new insights. Specifically, in this study, we use structural data from diffusion tractography recordings to model functional brain connectivity obtained from fast EEG dynamics occurring at the alpha frequency. First, we present a simple reference procedure which consists of several steps to link the structural to the functional empirical data. Second, we systematically compare several alternative methods along the modeling path in order to assess their impact

on the overall fit between simulations and empirical data. We explore preprocessing steps of the structural connectivity and different levels of complexity of the computational model. We highlight the importance of source reconstruction and compare commonly used source reconstruction algorithms and metrics to assess functional connectivity. Our results serve as an important orienting frame for the emerging field of brain network modeling.

Introduction

Resting-state brain activity represents the changes in neuroelectric or metabolic activity that occur when a subject is not performing a specific task and sensory input is largely reduced and stable. In this state spontaneous fluctuations emerge in the ongoing brain activity that synchronize across regions to exhibit a structured spatiotemporal pattern. Emerging resting-state networks have provided useful information regarding functional brain states, alterations in psychiatric or neurologic diseases, served as a basis for mapping and parceling the brain, and have helped to explain trial-to-trial fluctuations in cognitive functions (Fox et al., 2006; Arieli et al., 1996). Although electrophysiological recordings of brain activity have already revealed ongoing activity a long time ago (Steriade, Nunez, and Amzica, 1993; Contreras and Steriade, 1996; Destexhe, Contreras, and Steriade, 1999), the first description of common and organized networks emerging from ongoing activity was from functional Magnetic Resonance Imaging (fMRI)/Positron Emission Tomography (PET) studies which capture correlated slow fluctuations (< 0.1 Hz) across regions (Biswal et al., 1995; Raichle et al., 2001). Similarly, amplitude envelopes of alpha- and beta-frequency oscillations ($\sim 8-12$ Hz and $\sim 12-30$ Hz respectively) display similar correlation patterns as the fMRI signals and are usually oscillating at a similar slow time scale of < 0.1 Hz (Brookes et al., 2011; Hipp et al., 2012; Engel et al., 2013; Mantini et al., 2007). Both are here referred to as slow-fluctuating envelope resting-state networks.

The origin of resting-state ongoing brain activity is unresolved, but much evidence points to the anatomical skeleton shaping functional interactions between areas. A high dependency of slowly oscillating resting-state networks (< 0.1 Hz) and long-range axonal connections has been detected in several previous studies, indicating that local activity of segregated brain regions is

integrated by white matter pathways (Greicius et al., 2009; Honey et al., 2009; Skudlarski et al., 2008; Goñi et al., 2014; Hagmann et al., 2008). This structure-function relationship has also been explored in task-related functional networks and confirmed using different methodologies (Garcés et al., 2015; Chu et al., 2015; Vincent et al., 2007; Hermundstad et al., 2013). Although structural connectivity (SC) measured by diffusion tensor imaging (DTI) is seemingly a good predictor of functional connectivity (FC), functional connections also occur where there is little or no structural connectivity (Greicius et al., 2009; Honey et al., 2009). Honey et al. found that some of the variance in FC that could not be related to structure could, however, be accounted for by indirect connections and interregional distance (Honey et al., 2009). To explain missing links between anatomical structure and observed resting-state dynamics, bottom-up computational models based on structural priors offer interesting insights (Greicius et al., 2009; Honey et al., 2009; Skudlarski et al., 2008). Different computational models reflecting various biological mechanisms for the emergence of the spatiotemporal dynamics of resting-state networks have helped to explain the variance between SC and spatiotemporally organized low-frequency fluctuations (Cabral et al., 2014; Cabral et al., 2011; Hagmann et al., 2008; Izhikevich and Edelman, 2008). These dynamic simulations have robustly shown that the introduction of delays, scaling of coupling strength as well as additive noise lead to the emergence of functional patterns which resemble empirical resting-state networks operating in the low-frequency range. In this study, we first probe the assumption that computational modeling can also be successfully implemented to compare network simulations with empirical connectomes based on phase-relationships at fast frequencies. Secondly, we use our modeling framework to address several methodological issues of structural connectivity preprocessing, computational modeling and construction of functional connectomes.

Prior DTI-fMRI modeling studies have faced several technical challenges. First, the choice of computational model demands a trade off between highly simplified phenomenological models and biologically realistic models with a high dimensional parameter space. Surprisingly, as shown by Messé et. al (2014), a simple stationary model of functional connectivity better explains functional connectivity than more complex models (Messé et al., 2014; Messe, Benali, and Marrelec, 2015; Messé et al., 2015). Second, preprocessing of DTI data is necessary to derive a structural connectivity matrix on a given parcellation

scheme to overcome biases introduced by the latter. But the precise steps giving the most realistic structural connectome map are largely unknown.

Large-scale resting-state networks were originally described for correlated slow activity fluctuations recorded by fMRI/PET, or broadband power envelopes of the magneto-/electroencephalography (MEG/EEG) signal (Greicius et al., 2003). However, there is accumulating evidence that large-scale resting-state networks are also expressed in neuronal rhythms at faster frequencies (Mantini et al., 2007; Pasquale et al., 2010). Fast fluctuations in neuroelectric activity, and especially the functional linkage of regions via phase correlations, are well known to underlie a broad variety of cognitive processes (Singer, 1999; Fries, 2009; Engel and Fries, 2010; Siegel, Donner, and Engel, 2012). Synchronization of oscillatory neuronal activity among functionally specialized but widely distributed brain regions has been recognized as a major mechanism in the integration of sensory signals underlying perception and cognitive processes (Hipp, Engel, and Siegel, 2011; Hummel and Gerloff, 2005).

Regarding the spatial organization of fast oscillatory phase correlations, its quantitative relationship to SC has not been investigated yet (Deco, Jirsa, and McIntosh, 2011; Engel et al., 2013). Faster timescales of neural activity comprise for example the alpha, beta, or gamma band which constitute the major rhythms of spontaneous neuroelectric activity picked up by MEG/EEG. It has been argued, that compared to networks of slow fluctuations, structural connectivity does not strictly determine frequency-specific coupling in networks of ongoing activity at a faster timescale (Engel et al., 2013). Indeed, phase coupling between segregated areas strongly relies on cortico-cortical connections (Engel et al., 1991; Singer, 1999), implicating likewise a strong structure-function relationship.

Performance of the Reference Model In this study we probed this assumption of a strong structure-function relationship by simulating local node dynamics based on SC and comparing the phase relationships emerging from the simulated neural activity with empirically measured phase relationships. To this end, we combined SC from DTI data using probabilistic fiber tracking and FC from EEG data recorded during wakeful rest in 17 healthy individuals. We then used computational modeling approaches to link SC and empirical FC at

the alpha frequency range. We demonstrate that empirical networks of resting-state fast oscillations are strongly determined by the underlying SC and that additional variance between structure and function can be explained by modeling dynamic activity based on white matter architecture. Specifically, the simulated FC explained 28.5% of the variance in the empirical FC that was left unexplained by SC alone. To further understand the explanatory power of our model we investigated its performance at the local level by assessing specific properties of ROIs (nodes) or connections (edges). We found that the model error was highest for large highly interacting ROIs.

However, modeling large-scale brain dynamics based on structural priors brings up several methodological alternatives, not only regarding the modeling itself, but also regarding the comparison of simulated and empirical data. Especially with resting-state MEG/EEG activity, the specificity of analytic routines requires methodological decisions which potentially lead to tremendous differences in modeling outcomes. We systematically assessed the effect of technical variations on results and their influence on the interpretation of structure-function relations. Specifically, we used our modeling framework to explore several technical alternatives along the modeling path and evaluate the alternative processing steps based on their effect on the performance of the model in simulating empirical FC. Specifically, we addressed the effects of five critical aspects in the modeling pipeline:

Building the Structural Connectome We used DTI and probabilistic tracking algorithms to compile a whole-brain structural connectome (Bullmore and Sporns, 2009). However, several studies suggested that current fiber tracking algorithms fail at capturing particularly transcallosal motor connections that are observed in non-human primate tracer studies (Li et al., 2012a; Wedeen et al., 2008). In addition, structural connection strength modeled by probabilistic tractography algorithms is influenced by fiber length due to the progressive dispersion of uncertainty along the fiber tract (Li et al., 2012b; Goñi et al., 2014). Therefore, we evaluateed the effect of normalizations for fiber length of the SC and examined the effect of weighting homotopic connections in our model. Our results show that the correction for fiber distance leads to a small

decrease in the performance of our model. The additional weighting of homotopic transcallosal connections, however, increased the model fit (Messé et al., 2014; Messe, Benali, and Marrelec, 2015).

Model of Functional Connectivity Several alternative computational models of neural dynamics are available. In the choice of a more abstract version to a more realistic description of cortical interactions, these models vary in the complexity of their formulation and therefore might explain more or less variance in the observed FC. The downside of complex models, however, is the increased number of free parameters. These have to be approximated, need to be known a priori, or explored systematically. All these aproaches are problematic. For an assessment of the factor of model complexity, we compared a simple spatial autoregressive (SAR) model to the Kuramoto model of coupled oscillators. We find that the SAR model explains already a large portion of the variance and that the Kuramoto model only gives a slight improvement.

Forward and Inverse Models The comparatively few existing studies on large-scale modeling of MEG/EEG data differ systematically with respect to the comparison with empirical data. Some approaches project the observed time series onto the cortex using an inverse solution, whereas others project the simulated cortical signals into sensor space using the forward model (Cabral et al., 2014; Ritter et al., 2013; Bojak et al., 2010). We used our analytic framework to compare empirical and simulated FC at different spatial levels. We found that the importance of structural information is dramatically reduced if the higher spatial resolution obtained by source reconstruction is bypassed.

Source Reconstruction Algorithms Estimating the spatiotemporal dynamics of neuronal currents in source space generating the EEG and MEG signals is an ill-posed problem, due to the vastly larger number of active sources compared to the number of sensors. Therefore, we assess the impact of specific source reconstruction algorithms on the match of simulated and empirical FC. We compared three routinely used algorithms that differ regarding the assumptions made about the source signal, such as smoothness, sparsity, norms, correlation between source signals. However, we found no compelling superiority of one algorithm over another.

Functional Connectivity Metrics Functional connectivity describes statistical dependencies between two signals often based on undirected temporal averages such as correlation. In the last decades, various additional FC metrics have been introduced. These differ with regard to the relative weighting of phase and amplitude or concerning the removal of zero-phase lag components prior to correlation. The theoretical superiority of one approach over another is debated (Stam, Nolte, and Daffertshofer, 2007). However, no consensus appears achieved and currently no single metric is dominantly used over the others. Therefore, we compared several widely used metrics to compare empirical and simulated FC. We found that the model fit was much better if zero-phase lag components were preserved in the empirical functional connectome.

In the following sections, we first present a reference procedure for modeling FC based on DTI and the comparison with empirical FC as measured by EEG. After an initial short overview of the modeling approach (see the Workflow section), we guide the reader step by step through the model details with the resulting outputs of each processing stage (see the Reference Procedure section). From there, the impact of technical alternatives on the performance of the model is presented (see the Alternative Modeling Approaches section).

Results

Workflow

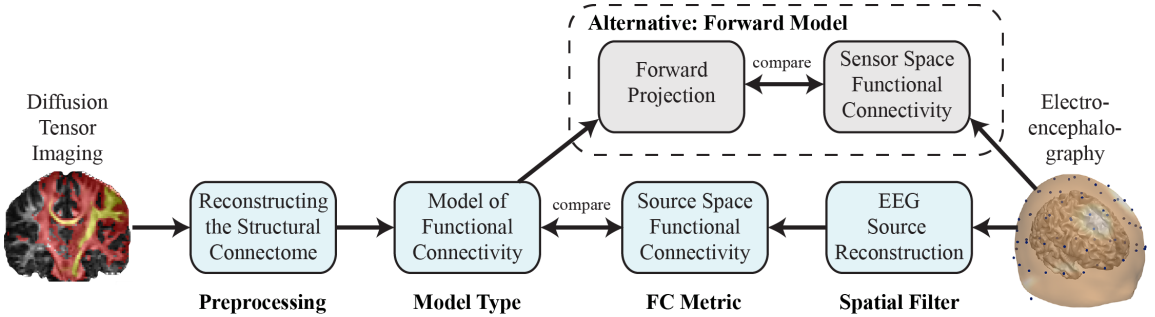
We compared the simulated FC based on SC with the empirical FC derived from EEG data (Fig. 2.1). Our model includes the processing steps as shown in Fig. 2.1 with the DTI measurements on the left and the EEG measurements on the right. We address preprocessing of DTI data in the form of homotopic reweighting. Then, the 66 ROIs of the cerebral cortex according to the 'Desikan-Killiany' cortical atlas made available in the Freesurfer toolbox, were individually registered for 17 healthy subjects using Freesurfer (surfer.nmr.mgh.harvard.edu) (Desikan et al., 2006). The SAR model used in the reference procedure was selected based on simplicity (low number of parameters) and performance (computationally very efficient). Furthermore, the SAR model allows to systematically evaluate the full parameter space with a high resolution grid-search, which is necessary for an unbiased comparison of

all alternatives along the modeling path. We reconstructed source activity at the geometric center of each ROI based on the EEG time series by a linear constraint minimum variance spatial beam former (LCMV). Then we assessed FC between source time series band pass filtered at 8 Hz where the averaged coherence showed a peak (see supporting material Figure S1 in Appendix A). Finally, we evaluated the match of simulated and empirical FC based on the correlation between all pairs of ROIs (Garcés et al., 2015). Following this modeling approach, several alternative ways at each processing stage arise. Choices exist, for example, for the level of abstraction of the model type (Sturm and König, 2001), metrics to compare functional connectivity and the approach to the inverse problem in interpreting EEG data.

Reference Procedure

Reconstructing the Structural Connectome The assessment of individual SCs is based on the number of probabilistic fibers connecting the parcellated brain regions. In our reference procedure, four preprocessing steps were applied to the raw fiber counts: First, we normalized the total number of tracked fibers between two regions by the product of the size of both regions. This effectively normalizes the connection strength per unit volume (Cheng et al., 2014). Second, we excluded all self-connections by setting the diagonal elements of the SC matrix (denoted as $\mathbf{S}$) to zero. The resulting SC matrix between the 66 anatomical ROIs is presented in Fig. 2.2A. Previous studies showed that current fiber tracking algorithms underestimate transcallosal connectivity (Li et al., 2012a; Wedeen et al., 2008). Accordingly, modeling studies have revealed that specifically increasing the SC between homotopic regions leads to a general improvement of the predictive power irrespective of the model (Messé et al., 2014; Messe, Benali, and Marrelec, 2015). Therefore, in the reference procedure we also increased the connection strength between homotopic regions by a fraction (h=0.1) of the original input strength at each node.

Last, we normalized the input strength of each region to 1, as done in previous simulation studies (Messé et al., 2014; Cabral et al., 2011). This normalization of the total input strength per region is based on the assumption that the DTI structural connectivity only informs about relative contributions to the input of each individual brain region. Or, stating it differently, DTI data does not



	Preprocessing	Model Type	FC Metric	Spatial Filter
Reference	Homotopic reweighting	SAR	Coherence	LCMV
Alternatives	Normalization with sum With distance correction No homotopic reweighting	Kuramoto Direct SC	Phase locking value Imaginary coherency (Weighted) phase lag index Lagged phase coherence	ELORETA MNE

FIGURE 2.1: **Workflow from DTI to the model of functional connectivity and comparison with empirical EEG data.** Each processing step in the reference procedure can be replaced by several alternative methods. From left to right: Probabilistic tracts derived from DTI are preprocessed to give the structural connectivity matrix. From there we simulate functional connectivity and optimize free model parameters to maximize the global correlation with the empirical functional connectivity. The empirical functional connectivity is calculated between all pairs of ROIs after projecting EEG scalp recordings to source space using spatial filters. Alternatively, the comparison between simulated and empirical connectomes can be done in sensor space by projecting the simulated functional connectivity into sensor space using the leadfields.

contain information about how much total input strength each individual region receives, but only relative input contributions per region.

Model of Functional Connectivity Several computational models of neural dynamics have been presented previously, varying in complexity regarding cellular and circuit properties (Cabral, Kringelbach, and Deco, 2014; Messé et al., 2014; Breakspear, Heitmann, and Daffertshofer, 2010). In the reference procedure, we chose a model of FC which is as simple as possible while still explaining a substantial fraction of the variance in the empirical data. For resting-state FC derived from fMRI data, it was shown that the simple SAR model generates good matches at low computational expense (Messe, Benali, and Marrelec, 2015; Messé et al., 2015). Therefore, we used the SAR model as a reference to evaluate just the static higher order dependencies in the FC.

The SAR model assumes that the time series of each region is a linear combination of the fluctuations of the time series of all other regions with added Gaussian noise, where only instantaneous effects are modelled. The activation of all ROIs $\vec{y}$ in the steady state condition is given by

$$\vec{y} = k \cdot \mathbf{S} \cdot \vec{y} + \sigma \cdot \vec{\nu}. \tag{2.1}$$

where $\mathbf{S}$ is set to the preprocessed SC matrix averaged across subjects as explained in the previous section. k is a global parameter describing the scaling of the coupling strengths. $\vec{\nu}$ is uncorrelated Gaussian noise that is added at each node individually and is scaled by σ. This equation describes the equilibrium state of the autoregressive model.

The covariance between the time series of the SAR model can be solved analytically by substituting (Tononi, Sporns, and Edelman, 1994)

$$\mathbf{Q} = (\mathbf{I} - k \cdot \mathbf{S})^{-1}, \tag{2.2}$$

so that

$$\vec{y} = \sigma(\mathbf{I} - k \cdot \mathbf{S})^{-1}\vec{\nu}. \tag{2.3}$$

The covariance matrix between sources is then given by

$$\mathbf{Cov} := < \vec{y} \cdot \vec{y}^T >_t = < (\sigma \cdot \mathbf{Q} \cdot \vec{\nu})(\sigma \cdot \mathbf{Q} \cdot \vec{\nu})^T >_t = \mathbf{Q} \cdot \mathbf{\Sigma} \cdot \mathbf{Q}^T, \tag{2.4}$$

where $<>_t$ denotes the average over time and $\mathbf{\Sigma} = \sigma^2 < \vec{\nu} \cdot \vec{\nu}^T >_t = \sigma^2 \mathbf{I}$ the noise covariance. Due to the assumption of uncorrelated Gaussian noise $\mathbf{\Sigma}$ is the identity matrix.

A FC matrix is constructed based on all pairwise correlations between network nodes. This can be calculated using the standard definition of correlation given the covariance from equation 2.4:

$$\mathbf{Corr}_{ij} = \frac{\mathbf{Cov}_{ij}}{\sqrt{\mathbf{Cov}_{ii} \cdot \mathbf{Cov}_{jj}}}. \tag{2.5}$$

This step normalizes for different variances in the time series of different network nodes. The resulting correlation matrix, as shown in Fig. 2.2B, is the predicted FC generated by the model given SC. The distribution of modeled FC is less sparse than the raw structural connection strength values: In the SC (Fig. 2.2A), many pairwise connections are close to zero and only few pairwise connections are large. To quantitatively evaluate the difference between the SC and the model output, we calculated the kurtosis of the values in the connectivity matrices:

$$Kurt[\mathbf{X}] := \frac{< \mathbf{X}_{ij}^4 >_{ij}}{< \mathbf{X}_{ij}^2 >_{ij}^2}, \tag{2.6}$$

where $<>_{ij}$ denotes the average over all upper triangular matrix elements without the diagonal (i.e. $i < j$). In this definition we divide the fourth raw moment by the second raw moment, where raw means that the moment is about the origin in contrast to central moments about the mean. The SC has a very high kurtosis ($Kurt[S] = 62.83$), whereas the FC predicted by the SAR model has a much smaller kurtosis ($Kurt[Corr] = 5.77$), indicating reduced sparsity.

Source Reconstruction Algorithms The spatiotemporal dynamics of neuronal currents in source space can be estimated using various source reconstruction techniques applied to the MEG/EEG signal. The algorithms differ regarding the assumptions made about the source signal (i.e. smoothness, sparsity, norms, correlation between source signals). These assumptions about the signals to be reconstructed are a prerequisite to make the ill-posed inverse

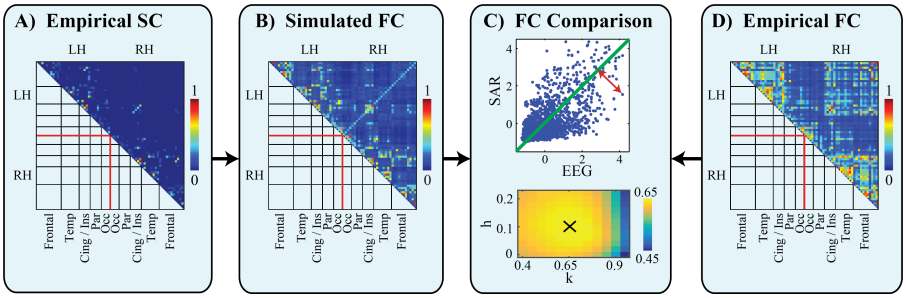


FIGURE 2.2: **Comparison of empirical and simulated FC in the reference procedure.** A: Structural connectivity among 66 cortical regions after normalization for ROI size and excluding self-connections (see chapter Reference Procedure, section Reconstructing the Structural Connectome). B: The correlation of the simulated network based on structural connectivity using the SAR model with optimal global scaling parameter k=0.65 and homotopic connection strength h=0.1. C: Upper: The respective simulated (k=0.65, h=0.1) and empirical connection strengths are z-transformed and plotted for each connection. Correlation is used as a global performance measure. The local model error per connection is evaluated as the distance (red arrow) to the total-least-squares fit (green line). Lower: Color indicates the correlation strength at a range of different global connection strength scaling parameters k, and fraction of added homotopic connections (h). The black cross indicates the parameters with the maximum correlation. D: The empirical functional connectivity as the coherence between source reconstructed time series at the cortical regions. All connectivity matrices (A,B,D) were normalized to have strengths between 0 (no connection) and 1 (strong connection).

problem of distributed sources treatable. As a reference, we used a LCMV spatial beamformer, which reconstructs activity with unit gain under the constraint of minimizing temporal correlations between sources (Van Veen et al., 1997). This approach has been applied in large-scale connectivity and global modeling studies before (Cabral et al., 2014; Garcés et al., 2015; Hindriks et al., 2015). Multichannel EEG data was projected to source locations based on individual head models. The spatial filter was calculated for the optimal dipole orientation corresponding to the direction of maximum power, thus giving one time series per ROI. As a priori source locations we used the geometric center

of each of the 66 ROIs individually registered on T1 images. See supplementary material (Empirical Data in Appendix A) for details on data acquisition, preprocessing and analysis of EEG data.

Functional Connectivity Metrics FC can be assessed using several methodologies which differ with regard to the relative weighting of phase and amplitude or concerning the reduction of zero-phase lag components prior to correlation (Vinck et al., 2011). The choice of metric may have an influence on the match between empirical and simulated FC. In the reference procedure, we calculated ordinary coherence as a metric for FC due to its original and prepotent implementation in synchronization studies (Rappelsberger, Pfurtscheller, and Filz, 1994; Andrew and Pfurtscheller, 1996; Gerloff et al., 2006; Manganotti et al., 1998; Hipp, Engel, and Siegel, 2011; Wu et al., 2014; Anguera et al., 2013; Niso et al., 2015). The time series at each source were bandpass filtered and then Hilbert transformed. Functional importance of resting state phase coupling networks at different frequencies has been demonstrated (Cabral et al., 2014; Hipp et al., 2012), motivating a correlation of simulated FC with empirical FC at different frequencies (see supporting material Figure S1 B in Appendix A). We found a comparably high model performance across several frequencies, highlighting that our main finding of simple computational models being able to explain missing variance between structure and function holds across several frequency bands. In the reference procedure we chose the alpha band since a broad spectrum (3-30 Hz) exploration showed a peak of the mean coherence across all connections at around 8 Hz (see supporting material Figure S1 in Appendix A).

The FC metrics are based on the analytic signal representation

$$A_m(t) = r_m(t) \cdot \exp(i\varphi_m(t)) \tag{2.7}$$

of region m. Furthermore, we calculated the cross-spectrum between two regions of interest m and n as

$$s_{m,n}(t) = A_m(t) \cdot \overline{A_n(t)}. \tag{2.8}$$

Given the analytic signal, the auto- and cross-spectra were computed and the

coherence derived as the normalization of the cross-spectrum by the two auto-spectra (Andrew and Pfurtscheller, 1996). This gives a FC index ranging from 0 to 1 between all pairs of ROIs:

$$\mathrm{Coh}_{m,n} = \Big\langle \frac{s_{m,n}(t)}{\sqrt{s_{m,m}(t) \cdot s_{n,n}(t)}} \Big\rangle_t. \tag{2.9}$$

The resulting mean empirical FC matrix across the group is depicted in Fig. 2.2D and was compared with the modeled FC matrix. Intrahemispherically, we found high connectivity within frontal and temporal areas in both hemispheres. Interhemispherically, the insular and cingulate areas were strongly connected.

Performance of the Reference Model The SAR model yields a FC of the 66 parcellated brain regions in accordance with the empirical FC. Since both these matrices are symmetric, only the triangular parts are compared to assess the match between simulated and empirical FC.

We calculate the performance of the model as the correlation between all modeled and empirical pairwise interactions (Fig. 2.2C). This performance metric is also commonly used in other studies (Messé et al., 2014; Cabral, Kringelbach, and Deco, 2014). We found a high correlation between the FC from the model and EEG coherence values (r=0.674, n=2145, $p < .0001$) for the parameters k=0.65 (global parameter describing the scaling of the coupling strengths) and h=0.1 (additional weighting of the homotopic connections in the SC matrix) marked in Fig. 2.2C below).

To put this into context, we first compared these results with the match between the empirical SC and FC without modeling (r=0.4833, n=2145, $p < .0001$) and found a shared variance of 23.4% (variance explained is $100 \cdot r^2$). Modeling FC based on this SC backbone increased the global correlation to 45.4% (square of r=0.674). In other words, the modeled FC explains roughly 28.8% of the variance in the empirical FC that is left unexplained by SC alone.

As a comparison to these results obtained from the average subject data, we also calculate the performance of the reference model based on the DTI and EEG data of individual subjects. The average correlation between modeled and empirical single-subject functional connectivity is (r=0.53408, n=2145, $p < .0001$) for matching DTI and EEG subjects. As a comparison, we evaluated

the performance when comparing nonmatching DTI and EEG subjects, which leads to a similar value (r=0.53362, n=2145, $p < .0001$). This small difference between matching and nonmatching subjects was statistically non-significant (p=0.48, tested using a linear mixed effects model), probably due to the low sample size and a low signal-to-noise ratio at the level of individual subjects.

To further understand the explanatory power of our model we investigate its performance at the local level by assessing specific properties of ROIs (nodes) or connections (edges). We defined for each connection the local model error as the distance (example shown as red arrow in Fig. 2.2C, upper) between each dot and the total-least-squares fit (green line in Fig. 2.2C, upper). Specifically, the question arises whether the high correlation between modeled and empirical FC is driven more by long or short edges. For example, the FC estimation between very close ROIs (in Euclidean space) might be spuriously inflated by volume conduction. Alternatively, there might be an overestimation of the SC between specifically close regions which could cause a higher model error (Jbabdi and Johansen-Berg, 2011). To address this question we compared for each edge the model error with the fiber distance (Fig. 2.3A). The average fiber distance between connected ROIs was negatively correlated with the logarithm of the local model error of each connection (r=-0.32, n=2145, $p < .0001$). A similar dependence was calculated between Euclidean distance between ROI locations and local model error (r=-0.33, n=2145, $p < .0001$). Both results indicate that the SAR model performed worse in simulating FC for closer ROIs in topographic space (measured in fiber lengths) and Euclidean space (measured as distance between ROI locations). This can be attributed to a higher variance in the SC and empirical FC matrices for close ROIs (as shown in supporting Figure S2 in Appendix A).

The empirical structural and functional connectivity are both dependent on the interregional distance between nodes with higher connectivity for short-range connections and lower connectivity for long-range connections (Bullmore and Sporns, 2012; Betzel et al., 2016). Therefore, we also calculate the model performance of our reference procedure after regressing out the distance between regions. The remaining partial correlation between modeled and empirical functional connectivity is $r = 0.36$ after regressing out the euclidean distance. A similar partial correlation $r = 0.38$ was calculated after removing the effect of fiber distance.

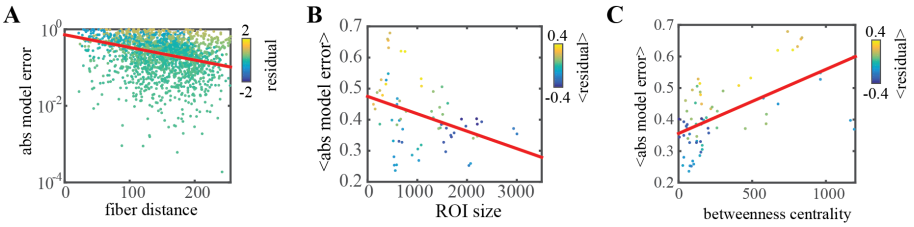


FIGURE 2.3: **Dependence of residual and model error (absolute value of residual) on edge and node characteristics.** A: linear fit of the log of the model error per connection showing a negative correlation with fiber distance. B: linear fit of the average model error per ROI showing a negative correlation with the size of the ROI. C: linear fit of the average model error per ROI showing a negative correlation with the betweenness centrality of the ROI. The angle brackets <> denote the average over all edges of the corresponding ROI. Residuals in A-C are calculated from the total least squares fit, negative values (blue dots) indicate that the average modeled functional connectivity per node was higher than the empirical functional connectivity, positive values (yellow dots) indicate that the the modeled functional connectivity per node was smaller than the empirical functional connectivity.

We further evaluated the performance in relation to certain node characteristics and averaged the errors of all edges per node. The node performance in terms of model error is shown in Fig. 2.3B-D dependent on different node characteristics. First, we looked at the influence of ROI size on the model error. We hypothesized that due to larger sample sizes and more precise localization, the model error would be smaller for large ROIs. As expected, the model error for each ROI is negatively correlated with the corresponding size of the ROI (r=-0.37, n=66, $p < .005$) as shown in Fig. 2.3B. Then we hypothesized, that due to the sparseness of SC, some ROIs in the SC have a very high connectedness compared to functional data, leading to a larger model error. To address this aspect we calculated several graph theoretical measures that assess the local connectedness in different ways and related this to the average model error. As a first measure we calculated for each node the betweenness centrality, defined as the fraction of all shortest paths in the network that pass through a given node (Rubinov and Sporns, 2010). The absolute model error is positively correlated with the betweenness centrality ($r = 0.58$, $n = 66$, $p < .0001$) as

shown in Fig. 2.3C. A similar indicator of a nodes connectedness in the network is the sum of all connection strengths of that node. Also for this metric, we find a linear relationship between the total connection strength of a node and the model error ($r = 0.35$, $n = 66$, $p < .005$). In addition, the dependence between the model error and the eigenvalue centrality, which measures how well a node is linked to other network nodes (Newman, 2008), was evaluated (r=0.26, n=66, $p < .05$). The local clustering coefficient, which quantifies how frequently the neighbors of one node are neighbors to each other (Watts and Strogatz, 1998), did not show significant relations with the local model error (r=0.06, n=66, $p = .65$).

Overall, the reference model can explain much of the variance in the empricial FC. The error in the predicted FC of the reference model appears to be highest for small highly interacting ROIs. This might be due to the more heterogeneous structure of small highly interacting ROIs. On the other side, interactions between more distant and large ROIs are better predicted by the model, probably due to the more homogenous connectivity.

Alternative Modeling Approaches

The modeling of large-scale brain dynamics based on structural priors brings up several methodological alternatives. As a principal choice, the model may be evaluated either in source or in sensor space. In the baseline model that was presented above, we made specific choices at each processing stage based on simplicity and good explanatory performance. Especially with resting-state EEG activity, a lack of analytic routines requires methodological decisions to be made heuristically, which could potentially lead to substantial differences in the conclusions drawn. In the following section we systematically compare different alternatives of the procedural stages delineated above and compare the outcome regarding global correlation between simulated and empirical FC. First, we assessed the influence of distance normalization and weighting of homotopic connections in the structural connectome on simulated FC. Second, we tested if a more complex simulation model of coupled oscillators is able to capture a larger part of the variance of the empirical data that is not explained by the simple SAR model. Third, we evaluated an alternative comparison in the sensor space using a forward projection of the source time series in contrast to source reconstruction. Then, we compared different source reconstruction

methods. Finally, we tested the impact of removing zero-phase lags in functional interactions.

Reconstructing the Structural Connectome The structural connectome was compiled using global probabilistic tractography. Interregional connections (edges) of the brain are represented by the number of "probabilistic streamlines" between these regions (nodes). We tested the performance of several alternative modifications of the SC (Fig. 2.4).

The pooled connectivity results obtained by the probabilistic fiber tracking are directly proportional to the size of the seed and target regions. The size of the regions, determined by the parcellation scheme, vary (Honey et al., 2009). They are parcellated based on standard gyral-based neuroanatomical regions (Desikan et al., 2006). In order to account for a bias of stronger connectivity of larger regions, SC was normalized using the size of the regions. However, the exact method of normalization for ROI size is currently a matter of debate and no operational routine has emerged yet (Cheng et al., 2012). Therefore, we compared different normalizations regarding the quality of the model. In the reference procedure, we normalized the number of tracked fibers between two regions by the product of the region sizes. We found that this approach (Fig. 2.4A) gives the best model performance (r=0.674, n=2145, $p < .0001$) in comparison with alternative normalizations that are presented in the following paragraphs.

First, instead of the normalization by the product of the two ROI sizes it is possible to normalize using the sum (Messé et al., 2014). However, the performance decrease in comparison to the reference procedure is very small ($r = 0.66$, $n = 2145$, $p < .0001$) as shown in Fig. 2.4B.

Second, an additional weighting was applied to correct for the influence of fiber length on the probabilistic tracking algorithm. Therefore, the streamlines connecting two regions were weighted by the corresponding fiber length. This normalization (Fig. 2.4C) leads to a small decrease in performance ($r = 0.65$, $n = 2145$, $p < .0001$).

Third, we tested the influence of homotopic transcallosal connections by omitting the additional weighting applied in the reference procedure. As a result,

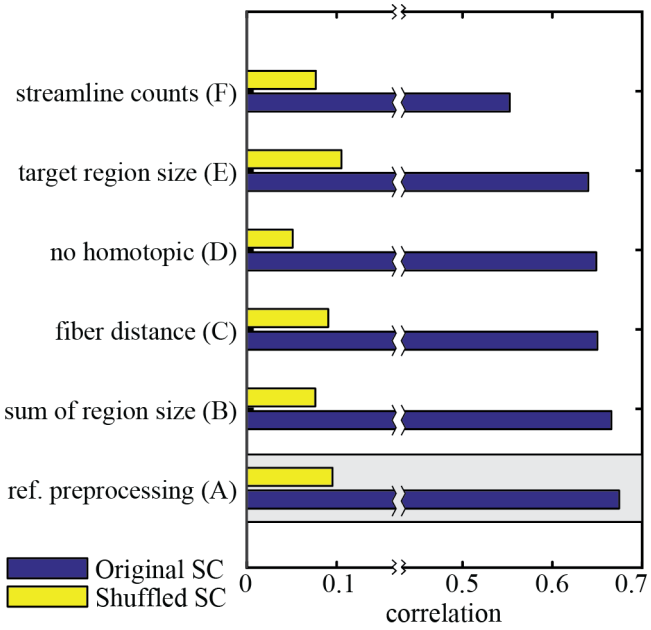


FIGURE 2.4: **Structural connectivity preprocessing.** The correlation between modeled and empirical functional connectivity for different preprocessing steps of structural connectivity. In the reference procedure, the number of tracked fibers between two regions was normalized by the product of the region sizes. The model based on the original structural connectivity is shown in blue and the baseline model which is based on shuffled structural connectivity in yellow. The gray box marks the reference procedure.

the correlation between modeled and empirical FC drops from $r = 0.674$ to $r = 0.65$ (Fig. 2.4D).

As a fourth alternative, we replaced the normalization by the product of region sizes by a normalization just by the target region in the simulation model (Cabral et al., 2011). This leads to a further small reduction of the performance to $r = 0.64$ (Fig. 2.4E).

As a last alternative we also evaluate the performance using just the normalized streamline counts as input to the model without any further preprocessing (no additional homotopic weights and no input strength normalization per region).

This baseline without further preprocessing has a lower performance with a correlation of $r = 0.55$ (Fig. 2.4F), suggesting that the normalization of the total input strength per node plays an important role for a good match with the empirical data.

These results demonstrate that our reference method of reconstructing the SC is superior to all the evaluated alternative approaches. Overall, the performance of the simulation based on the SC is rather robust with respect to the choices of preprocessing as long as the total input strength per region is normalized.

Model of Functional Connectivity In the previous sections we showed that a considerable amount of variance in empirical FC can be explained even with a simple SAR model that captures only stationary dynamics. Several alternative computational models of neural dynamics have been presented that vary regarding their complexity. More complex models can incorporate aspects of cortical processing at the microscopic scale such as cellular subpopulations with differing membrane characteristics or, at the macroscopic scale, time delays between nodes (Sturm and König, 2001; Cabral, Kringelbach, and Deco, 2014; Leon et al., 2013). The downside of complex models is the increased number of free parameters whose values need to be approximated, have to be known a priori, or explored systematically. We hypothesized that a more complex model which incorporates more parameters in order to simulate neural dynamics more realistically might explain more variance in FC. We decided to use the Kuramoto model of coupled oscillators as an alternative to investigate whether this holds true (Kuramoto, 1984; Finger and König, 2013; Cabral et al., 2011). In contrast to the SAR model, the Kuramoto model can incorporate delays between nodes and thus becomes a model of dynamic neural processes (Yeung and Strogatz, 1999; Breakspear, Heitmann, and Daffertshofer, 2010). At the same time the Kuramoto model is simple enough to systematically explore the parameter space. The progression of the phase of each neuron is modeled by the differential equation

$$\frac{\partial \varphi_j(t)}{\partial t} = 2\pi\omega - k \sum_{i \neq j} S_{ij} \cdot \sin\Big(\varphi_j(t) - \varphi_i(t - d - D_{ij}/\mathrm{v})\Big), \qquad (2.10)$$

where d is a fixed delay at each node and v is the transmission velocity which is weighted by the distance D_{ij} (see Empirical Data), which leads to a connection-specific delay. The Kuramoto model was simulated using the Euler integration method in time steps of 0.1 ms. In contrast to the SAR model, which does not reflect temporal dynamics, in the Kuramoto model we used the same bandpass filters and coherence estimation method as described in equations 2.7, 2.8, and 2.9.

An additional alternative to the SAR model is an even more simple direct comparison between the empirical SC and FC. The simple structure-function comparison gave a 23.4% match between structural and functional connectivity alone (r=0.4833, n=2145, $p < .0001$). The SAR model and the Kuramoto model both explain more variance of the functional connectivity than this direct comparison of structural and functional connectivity (Fig. 2.5A). Using the SAR model we simulated a functional connectome with a 45.4% match to the empirical data (r=0.674, n=2145, $p < .0001$). With the Kuramoto model however, the match could be further increased to 54.0% (r=0.735, n=2145, $p < .0001$). In other words, the modeled FC using the Kuramoto model explains 40.0% of the variance in the empirical functional connectivity that is unexplained by structure alone. In addition, demonstrating the importance of the underlying structural network, all three variants have a significantly higher correlation than for the randomly shuffled SC.

The Kuramoto model showed the best performance for a connection strength scaling of $k = 700$ (Fig. 2.5B). Important to note is that the constant delay can be neglected without a large performance drop (Fig. 2.5C). In contrast, the velocity introduces a connection specific delay that is modulated by the DTI fiber lengths and the model performance has a considerable peak around $v \approx 1.7$ m/s.

Forward and Inverse Models In the comparatively few studies on large-scale modeling of MEG/EEG data, a discrepancy exists to whether simulations are compared with empirical data in the source or sensor space (Cabral et al., 2014; Ritter et al., 2013; Bojak et al., 2010). In other words, the measured time series are either projected onto the cortex using an inverse solution or the simulated cortical signals are projected into sensor space using a forward model. Here we compare both approaches, source reconstruction vs. forward

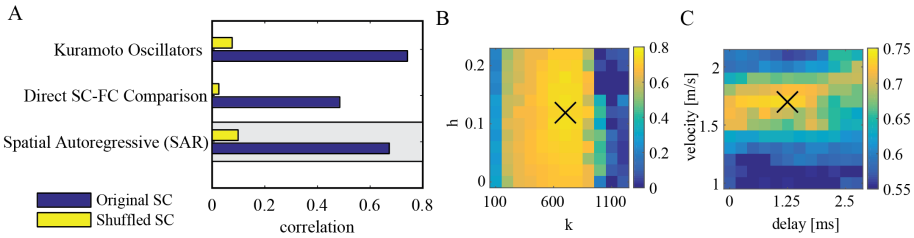


FIGURE 2.5: **Model of functional connectivity.** A: Performance comparison between the SAR model (reference model), the Kuramoto model and directly between the empirical and structural connectivity. The model based on the original structural connectivity is shown in blue and the baseline model which is based on shuffled structural connectivity in yellow. The gray box marks the reference procedure based on the SAR model. B: Performance of the Kuramoto model for different parameters k and h close to the optimal point with fixed $velocity = 1.7$ m/s and $delay = 1.25$ ms. C: Same as B but with varying velocity v and delay d with fixed $k = 700$ and $h = 0.12$. In panels B and C the X marks the parameter that was selected for the corresponding other panel.

projection, with respect to the global correlation strength between modeled and empirical FC. The source reconstruction approach has been described above (see chapter Source Reconstruction Algorithms and Empirical Data).

For the inverse solution and forward projection, we computed as a forward model a boundary element method volume conduction model based on individual T1-weighted structural MRI of the whole brain and comprising 8196 dipoles distributed over 66 regions (Oostenveld et al., 2003). Each dipole has six degrees of freedom defining its position, orientation, and strength in the cortex. The positions for each vertex are defined to be lying equally spaced within the parcellated brain regions of the cortical sheet. The electric source activity can be approximated by the fluctuation of equivalent current dipoles generated by excitatory neurons that have dendritic trees oriented perpendicular to the cortical surface (Ritter et al., 2013). For the inverse solution, the dipoles orientation was assessed according to its maximal power. For the forward projection of simulated time series, the dipole orientations were defined by the normal vector of the cortical surface of the corresponding region in the segmented MRI image. Since each of the parcellated brain regions extends

over several surface vertices, all dipole normals within each region are averaged. This results in one average direction vector per region (average length over all regions: 0.52) which is used to project into the EEG sensor space.

In the previous sections we showed that the underlying SC had a large impact on the relatively good match between simulated and empirical FC. Fig. 2.4 and Fig. 2.5A show large drops in correlation when the simulation is based on shuffled SC (yellow bars) instead of the original SC (blue bars). By comparing the source reconstruction with the forward model approach, we find that the comparison in sensor space using the forward projection yields higher correlations between simulated and empirical data (Fig. 2.6A). If, however, the underlying structural connectivity is shuffled before applying the SAR model, the correlation of simulated and empirical FC remains equally high in sensor space. This indicates that the importance of structural information is dramatically reduced if the higher spatial resolution obtained by source reconstruction is bypassed. The forward projection of the simulated time series leads to a very low spatial specificity of the functional connectivities in sensor space (Fig. 2.6B).

Since several inverse methods are routinely used without a clear superiority of one over another, we aimed to assess the impact of the specific source reconstruction algorithm on the fit between simulated and empirical FC. We compared three prominent and widely used inverse methods which make fundamentally different assumptions (Fig. 2.7). As a reference, we used an LCMV spatial beamformer which reconstructs activity with the constraint of minimizing temporal correlations between sources (Van Veen et al., 1997). For comparison we calculated the inverse solution by using exact low resolution brain electromagnetic tomography (ELORETA) which reconstructs activity by spatial smoothness constraints and in this sense it emphasizes local temporal correlations in comparison to beamforming approaches (Pascual-Marqui et al., 2011). It is also widely used in source connectivity analyses (Steinmann et al., 2014; Vecchio et al., 2015). Additionally we calculated the minimum-norm estimate (MNE) which recovers source activity by reducing overall energy (Dale et al., 2000) which is based on the assumption that the data gives no information about the null space component of the leadfield which is thus set to zero. Fig. 2.7 shows the global correlation values resulting from these three alternative inverse solutions. It can be seen that all of them have a similar performance level (LCMV:r=0.674, n=2145, $p < .0001$), ELORETA: (r=0.728,

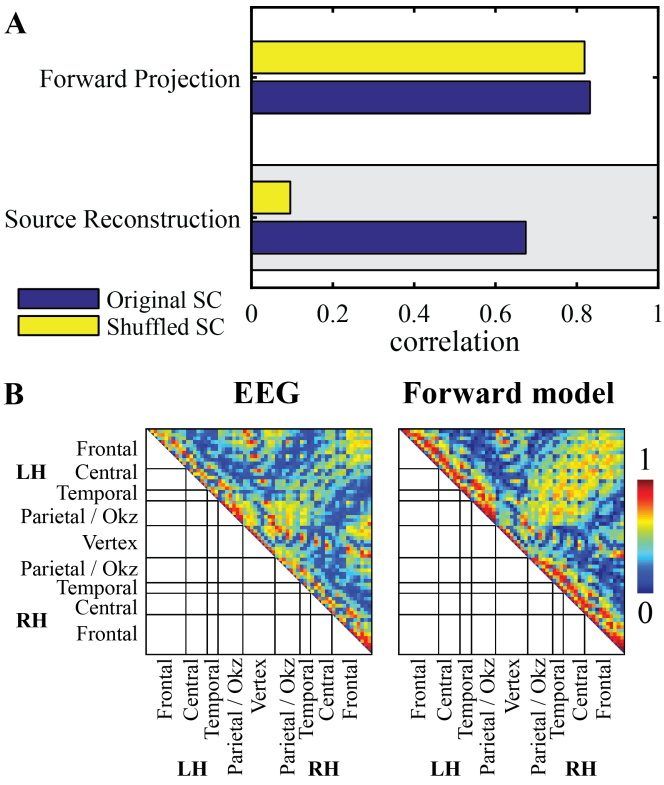


FIGURE 2.6: **Comparisons of forward projection and source reconstruction.** A: Global correlation between simulated and empirical functional connectivity in sensor space by applying the forward projection to the SAR model, or in source space by applying the LCMV beamformer to the EEG time series. Blue bars show simulations based on original structural connectivity and yellow bars simulations for randomly shuffled structural connectivity. The gray box marks the reference procedure. B: EEG functional connectivity measured by coherence (left) and the forward projected modeled functional connectivity (right), both in sensor space.

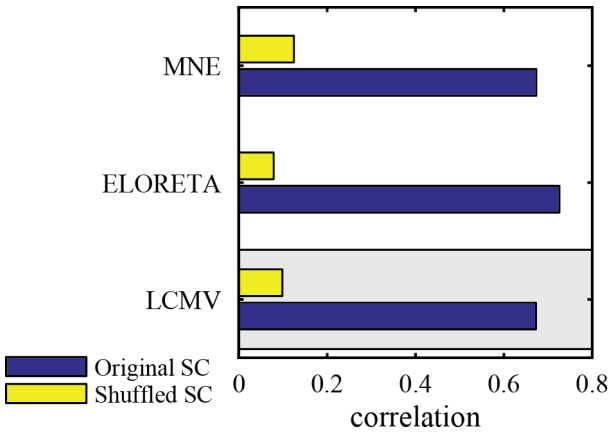


FIGURE 2.7: **Source reconstruction.** The correlation between modeled and empirical functional connectivity for different source reconstruction algorithms. The model based on the original structural connectivity is shown in blue and the baseline model which is based on shuffled structural connectivity in yellow. The gray box marks the reference procedure.

n=2145, $p < .0001$), MNE: (r=0.676, n=2145, $p < .0001$). The connectivity matrices of time series of the inverse solutions were highly correlated (LCMV-ELORETA: r=0.84, LCMV-MNE: r=0.95,MNE-ELORTEA: r=0.84; all $p < .0001$).

Functional Connectivity Metrics We compared several widely used FC metrics regarding the global relation between empirical and simulated functional connectivity. Previous modeling studies implemented different metrics, and clear superiority of one over another has not been shown (Lachaux et al., 1999; Stam, Nolte, and Daffertshofer, 2007; Vinck et al., 2011). In the reference procedure, empirical FC was calculated as ordinary coherence and compared to the FC matrix derived from the SAR model. In addition, we investigated several alternative FC metrics (Andrew and Pfurtscheller, 1996; Lachaux et al., 1999; Nolte et al., 2004; Stam, Nolte, and Daffertshofer, 2007; Vinck et al., 2011; Pascual-Marqui, 2007b).

TABLE 2.1: **Functional Connectivity Metrics**

Name	Abbreviation	Equation	Characteristic	Ref	Result
coherence	COH	$\text{Coh}_{m,n} = \left\lvert \left\langle \frac{s_{m,n}(t)}{\sqrt{s_{m,m}(t) \cdot s_{n,n}(t)}} \right\rangle_t \right\rvert$	Normalization of the cross-spectrum by the two auto-spectra	(Andrew and Pfurtscheller, 1996)	(r=0.674, n=2145, $p < .0001$)
phase locking value	PLV	$\text{PLV}_{mn} = \left\lvert \left\langle \mathrm{e}^{i \cdot \left(\varphi_m(t) - \varphi_n(t)\right)} \right\rangle_t \right\rvert$	Relative stability of the relative phase between two signals and thereby avoiding amplitude as a confounding factor	(Lachaux et al., 1999)	(r=0.678, n=2145, $p < .0001$)
imaginary part of the coherency	ICOH	$\text{Icoh}_{m,n} = \left\lvert \Im\left(\left\langle \frac{s_{m,n}(t)}{\sqrt{s_{m,m}(t) \cdot s_{n,n}(t)}} \right\rangle_t \right) \right\rvert$	Discarding the real part of the cross-spectrum, in which zero-phase lag volume conduction artifacts are exclusively represented	(Nolte et al., 2004)	(r=0.103, n=2145, $p = .37$)
phase lag index	PLI	$\text{PLI}_{m,n} = \left\lvert \left\langle \operatorname{sgn}\left(\Im(s_{m,n}(t))\right) \right\rangle_t \right\rvert$	Completely robust to volume conduction but without incorporation of the amplitude of the signal	(Stam, Nolte, and Daffertshofer, 2007)	(r=0.138, n=2145, $p < .05$)
weighted phase lag index	WPLI	$\text{wPLI}_{m,n} = \frac{\left\lvert \left\langle \Im(s_{m,n}(t)) \right\rangle_t \right\rvert}{\left\langle \lvert \Im(s_{m,n}(t)) \rvert \right\rangle_t} = \frac{\left\lvert \left\langle \lvert \Im(s_{m,n}(t)) \rvert \cdot \operatorname{sgn}(\Im(s_{m,n}(t))) \right\rangle_t \right\rvert}{\left\langle \lvert \Im(s_{m,n}(t)) \rvert \right\rangle_t}$	Similar to PLI, but more robust to noise in case of weak synchronization	(Vinck et al., 2011)	(r=0.170, n=2145, $p < .0001$)
lagged phase coherence	LPC	$\text{LPC}_{m,n} = \frac{\Im\left(s_{m,n}(t)\right)}{\sqrt{1 - \Re\left(s_{m,m}(t)\right)^2}}$	Zero-lag contribution partialled out by a quotient of residual variances conditional on the real and imaginary part or on the real part only; preserves parts of the real valued coherency which incorporate a phase lag	(Pascual-Marqui, 2007b)	(r=0.179, n=2145, $p < .0001$)

$\Re$ and $\Im$ denote the real and imaginary part of the complex number.

All metrics were based on the same analytic signal representation as shown in equation 2.7 and the cross-spectrum as defined in equation 2.8. The different metrics are listed in Table 2.1 with their corresponding equations, characteristics and results. Comparing the performances based on all five measures (see Fig. 2.8), we found a high correspondence in model performance between coherence and PLV. In contrast, PLI, WPLI, and LPC all showed a significantly lower match between simulated and empirical FC, with correlation coefficients between 0.10 and 0.18. ICOH showed the smallest correlation between modeled and empirical data with a non-significant p-value (r=0.103, n=2145, p=.37). For all metrics, the global correlation essentially vanished if the underlying SC was shuffled prior to simulation.

In summary, there are substantial decisions to be made at each stage of the processing pipeline. We selected the reference procedure prior to the evaluation of all alternatives in all these processing stages. Then, we evaluated the combination of all best performing alternatives along the pipeline. This best performing combination consists of the reference preprocessing of DTI data to construct SC, the Kuramoto oscillator network to simulate FC, PLV as a FC metric, and ELORETA as source reconstruction method from EEG. This combination results in a match of 54.4% between simulated and empirical functional connectivity (r=0.7377, n=2145, $p < .0001$).

Discussion

With this study we contribute to resolving the structure-function relationship in global connectomics. We simulated fast neural dynamics based on a realistic structural connectivity backbone and compared it to empirical functional connectivity derived from phase coupling of oscillatory brain waves. For the empirical data collected in 17 subjects we found a 23.4% match between structural and functional connectivity alone. Using a simple SAR model to simulate FC based on SC, this match was increased to 45.4%, showing that the model can capture about 28.8% of the variance in the empirical FC that is unexplained by the structure alone. We demonstrate several technical alternatives in the modeling procedure and derivation of empirical connectomes that are commonly used, but only few gave noticeable improvements. Of note, introducing

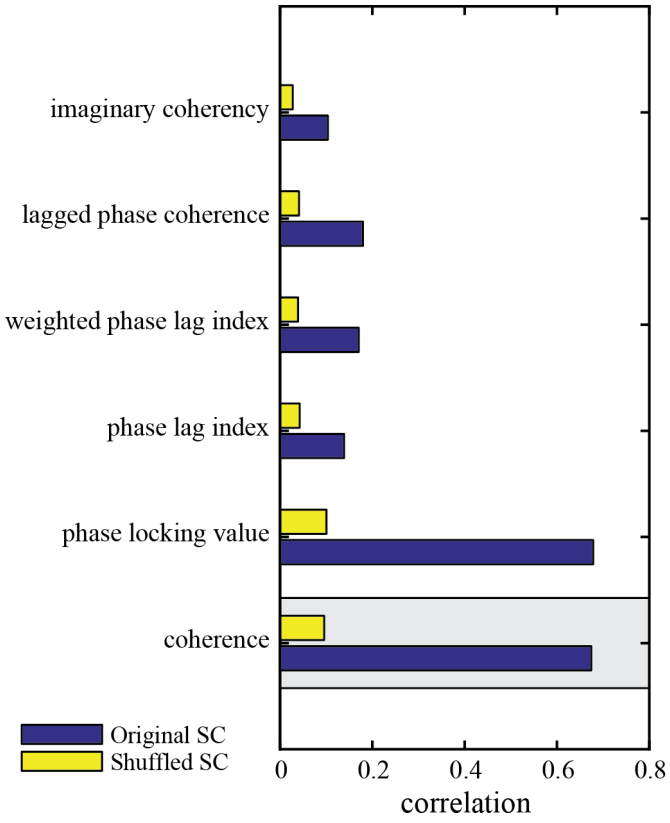


FIGURE 2.8: **Functional connectivity metrics.** The bars show the correlation between the empirical functional connectivity and the simulated functional connectivity obtained using the SAR model. The model based on the original structural connectivity is shown in blue and the baseline model which is based on shuffled structural connectivity in yellow. The gray box marks the reference procedure.

additional model parameters by using the Kuramoto model of coupled oscillators improved the simulation (Fig. 2.5). Our results show that resting-state networks emerging from phase coupling at a fast timescale largely resembles structural connectivity, as it has been previously shown for slow fluctuations of BOLD-signal or broad-band power envelopes (Hagmann et al., 2008; Greicius et al., 2009; Honey et al., 2009; Cabral et al., 2014).

Modeling Fast Dynamics

It has been assumed that for the resting-state networks based on fast dynamics the underlying anatomical skeleton is less important compared to the slow resting-state networks, but this issue has not yet been systematically investigated (Engel et al., 2013). We calculated the performance of the reference model as the correlation between all modeled pairwise interactions and all empirical pairwise interactions in an empirical functional phase relation connectome of the alpha rhythm and found a good match of 45.4% (Fig. 2.2C). This finding is in contrast to the prior assumptions and shows that the anatomical skeleton is equally crucial for fast timescale functional interactions (Engel et al., 1991; Singer, 1999).

To better understand the reference model performance we investigated the model error in relation to node and edge characteristics (Fig. 2.3). In general, the model error decreased with longer fiber distance and Euclidean distance. Specifically, for short fiber distances, the model overestimated FC (negative residuals blue in Fig. 2.3). Why are short connections in general more difficult to model based on white matter tracts? The empirical connectome was extracted from resting-state alpha topographies in which propagating waves play an important role for adjacent and remote brain areas to communicate with each other. Cortico-cortical axons in the white matter tracts are considered as the major route for traveling waves. However, a recent study presented compelling evidence for intracortical axons accounting for spatial propagation of alpha oscillations (Hindriks, Putten, and Deco, 2014). Such a mechanism would enable high local synchrony in the relative absence of structural connectivity measured by DTI.

We used a stepwise linear model to extract node characteristics explaining most of the model error and found that ROI size and betweenness centrality

play an important role. Regarding ROI size, the smaller model error for larger ROIs could be attributed to the measurements of structural as well as functional connectivity being more reliable for larger ROIs: In the case of the SC measurements using DTI, a larger parcellated cortical region allows to track more streamlines with different initial conditions (i.e. for more voxels) and thereby allows a more reliable estimation of the connection probabilities between regions. In EEG as well as DTI, the localization and inter-subject registration of large ROIs can be assumed to be less effected by small deviations because a small spatial shift of a large ROI still allows a large overlap with the correct ROI volume whereas a small spatial shift of a small ROI could displace it completely outside of the original volume. For betweenness centrality, the opposite scenario was the case: the smaller the betweenness centrality the smaller was the model error. Central hubs in a structural network offer anatomical bridges which enable functional links between regions that are structurally not directly related (Rubinov and Sporns, 2010). Hard-wired connections do not necessarily contribute at all times to FC in the network and, vice-versa, functionally relevant connections do not necessarily have to be strongly hard-wired (Honey et al., 2009). Possibly, the simple SAR model, which captures only stationary dynamics, has weaknesses at these central hub nodes. In order to capture the empirical FC at these nodes, a more complex dynamical model able to capture non-stationary dynamics with context switches at slower time scales is needed. Nodes with a high betweenness centrality can be expected to communicate with certain cortical modules only at certain times in specific dynamical regimes. We hypothesize that a more complex dynamical model of neural activity could capture this behavior more accurately. Therefore we suggest that further research could especially improve the model in these cases of dynamical context switches in central hub nodes, which cannot be captured by simple models such as the SAR model.

Reconstructing the Structural Connectome

Using our modeling framework to compare different alternatives of reconstructing the structural connectome, we found that the best match between simulated and empirical FC was obtained when an additional weighting of connections between homotopic transcallosal regions was applied. Additional

weighting for fiber distances did not improve the simulation performance significantly. Overall, the differences were very small proving the modeling approach to be rather robust regarding the evaluated choices of reconstruction as long as the total input strength per region is normalization prior to the simulation.

Currently, there is no common approach to correct for the influence of fiber distance on the probabilistic tracking algorithm (Hagmann et al., 2008; Supekar et al., 2010; Li et al., 2012b). Although we found that the model error was largest for small fiber distances (modeled FC higher than empirical FC), a correction for fiber lengths did not improve the result of the simulation. This suggests that the high local connection strength of SC obtained by DTI reflects actual structural connectivity. Methodically, this finding is supported by the fact that accuracy of probabilistic fiber reconstrunction decreases with distance, whereas short-distance connections are reconstructed with high reliability (Li et al., 2012a). However, it remains a challenge to correct probabilistic tracking results for the impact of fiber distance and further work is needed to address this methodological limitation.

Our model improved with an additional added weight of homotopic connections, which is supporting the data by Messé et al. (Messé et al., 2014). This finding points to a related limitation of the probabilistic tracking algorithms to correctly assess long distance and lateral transcallosal fibers. In agreement with previous studies, we show that this limitation can be addressed by adding a preprocessing step to the structural connectome reconstruction.

Lastly, we want to point out that the parcellation scheme, especially the spatial resolution, has a strong effect on the SC and FC, as shown in previous studies (Zalesky et al., 2010; Zhong, He, and Gong, 2015; Power et al., 2011). We did not include other parcellation schemes as alternatives in this work because a different parcellation effects all steps in the processing pipeline at the same time. Most importantly, a different parcellation also changes the predefined space in which the model prediction is evaluated, so that the resulting correlation values are not directly comparable to the results of our presented reference procedure. We chose a parcellation scheme which has been used in several previous studies (Hagmann et al., 2008; Li et al., 2012b; Desikan et al., 2006; Ton, Deco, and Daffertshofer, 2014; Cabral et al., 2011) and implemented in Freesurfer. The effect of parcellation schemes on structure-function

relationships is a very important topic that is currently under investigation.

Model of Neural Activity

We show that our SAR model already explains much of the variance in the empirical EEG data. Our results indicate that the Kuramoto model moderately improved results compared to the reference model. The SAR model has a small number of parameters allowing a fast exploration of the parameter space (Tononi, Sporns, and Edelman, 1994) and the SAR model served several studies in which complexity and information-theoretical measures characterizing FC were explored (Tononi, Sporns, and Edelman, 1994; Sporns, Tononi, and Edelman, 2000; Barnett, Buckley, and Bullock, 2009). As a downside, the SAR model has a smaller number of parameters and therefore lacks the modeling capacity to further optimize the dynamics to better fit to the empirical data. Furthermore, the SAR model cannot model individual frequencies and their interactions, making the Kuramoto model a viable alternative. It has been shown that the Kuramoto model features complex synchronization dynamics which can be related to the explanation of oscillatory phenomena in the human cortex, such as fluctuating beta oscillations (Breakspear, Heitmann, and Daffertshofer, 2010) or metastable synchronization states (Cabral et al., 2014). A more detailed analysis of the synchronization properties of the Kuramoto model in the human connectome was done by Villegas et al. (Villegas, Moretti, and Muñoz, 2014), where frustration and the transition between synchronous and asynchronous phases were analyzed (Sadilek and Thurner, 2015). The Kuramoto model was also used to study the effects of lesions on cortical dynamics and binding by synchrony (Váša et al., 2015; Finger and König, 2013). However, it has been shown that more complex models with more parameters are usually not better in explaining fMRI functional connectivity from structural data (Messé et al., 2014; Messe, Benali, and Marrelec, 2015; Messé et al., 2015). Highly parameterized models which require the numerical integration of differential equations take several orders of magnitude more computational time to obtain a reliable estimate of FC than the simple model used here. For certain neurophysiological questions however, the wider parameter space of complex models can be used to explore neural processing properties. The relative benefit of a dynamical model has to counterbalance the higher computational demand. Therefore, the choice of model depends on the

investigated scientific question (Sturm and König, 2001; Messé et al., 2015). In this study we used the simpler SAR model as a reference because the focus was to investigate alternatives also in many other stages of the processing pipeline and a more complex simulation model would impede identifying the best alternative in the other stages of the processing pipeline, due to the high dimensional parameter space.

Source Reconstruction versus Forward Projection

The inverse problem is ill-posed since the higher number of possible active neuronal sources is higher than the number of recording channels. Thus, the ground truth of brain activity patterns generating the measured signal is impossible to infer. A variety of alternative methodological approaches have been developed regarding source imaging. Particular caution should be exercised concerning the influence of different inverse solutions on the resulting data (Schirner et al., 2015). Here, we presented a comparison of the performance of three commonly used inverse methods regarding the global correlation between empirical and simulated FC in our technical framework (Fig. 2.7). All source reconstruction algorithms perform in a similar range with resembling r-values between 0.674 and 0.728. Although the algorithms differ regarding the assumptions made about the source signal, the high correspondence in performance of the three source reconstruction techniques mutually validates their respective inverse solutions.

Next, we aimed to investigate the best approach for comparing empirical and simulated FC particularly in sensor and source space, see Fig. 2.1. In the sensor space scenario, the simulated signal, as the mean field source activity generated by the SAR model, was projected into sensor space to generate a simulated EEG signal by applying the leadfield (i.e. forward model). For this approach we found slightly higher correlations between simulated and empirical data (Fig. 2.6A). However, we also found that the high correspondence between empirical and simulated EEG sensor space FC was independent of the underlying SC: Shuffling SC before the simulation did not abolish the correlation between the empirical and simulated FC as was the case when the comparison was done at the source space level. This lack of specificity of the simulated FC regarding the anatomical skeleton strongly suggests that the sensor level connectivity matrix is shaped mainly by the leadfield (Fig. 2.6). In fact, the

leadfield can already explain most of the variance (81.9%) in the empirical FC of the sensor space. In contrast, the inverse solution in the source reconstruction procedure removes much of these volume conduction correlations so that the comparison of coherence in source space appears reasonable. We conclude that the volume conduction model of the head is mixing the source time series such that the coherence in sensor space reflects to a high degree the structure within this mixing matrix and the sensor space is a suboptimal stage for investigating structure-function relationships by large-scale modeling approaches. Thus, one should refrain from such a comparison in sensor space with metrics that do not exclude zero-lag interactions. In order to assess the accuracy of simulated global network characteristics, the comparative spatial level should be at source space in order to avoid signal mixing by the leadfield matrix and allow to include zero-lag interactions. The results offer an important ground for modeling studies using source connectivity analyses for MEG/EEG data.

Connectivity Metrics and the Role of Phase Lags

One of the main differences between fMRI/PET and MEG/EEG connectivity studies is that for MEG/EEG a multitude of different metrics to quantify FC are currently available and no single metric is predominantly employed or has emerged as being superior (Vindiola et al., 2014; Greenblatt, Pflieger, and Ossadtchi, 2012). This issue hampers comparability between studies and physiologic interpretation. It was our aim to use our theoretic framework for a systematic comparison of different functional connectivity metrics. We compared six commonly used metrics that differ regarding their sensitivity towards zero-phase lag coupling and amplitude variations. The definition of PLV, PLI, WPLI and LPC theoretically renders those metrics insensitive to amplitude variations. We found no major difference in performance between COH and PLV and no major difference between ICOH, PLI, WPLI, and LPC. This result is easily understood on the basis that the SAR model presents the steady state solution including a small noise component only. An important finding is the high correspondence in model performance between coherence and PLV. Coherence is the cross-spectrum between two sensors normalized with the autospectra whereas PLV quantifies the consistency of a phase difference between two signals across time. Both measures are high if there is a consistent phase difference regardless of whether the latter is near zero, 180° or inbetween.

Similar results between coherence and PLV have been found in previous studies (Krusienski, McFarland, and Wolpaw, 2012; Khadem and Hossein-Zadeh, 2014; Zhang et al., 2014). The similarity of both measures in our study suggests that amplitude variations between areas are of less weight than phase variations. Another main finding is the drop in model performance with the metrics ICOH, PLI, WPLI and LPC which are by design less sensitive to zero-phase coupling. Regarding the latter, a major concern exists whether such coupling in scalp recordings would be contaminated by volume conduction artifacts. Obviously, synchrony at sensor level could result from two channels picking up activity from a common source since the activity of the source signal passes through the layers of cerebrospinal fluid, dura, scalp and skull acting as a spatial filter. This effect leads to the detection of spurious synchrony, even if the underlying sources are independent (Nunez and Srinivasan, 2006). Based on the assumption that the quasi-static approximation holds true for EEG, volume conduction would occur with zero-phase lag (Stinstra and Peters, 1998). Thus, the most commonly used approach to deal with the problem of volume conduction is to neglect interactions that have no phase delay. This is, however, a potentially overly conservative approach. To address the question of how these biased measures of interactions are suited for comparing modeled and empirical connectomes, we compared global model performance based on connectivity metrics that are sensitive and robust to zero-phase lags in this study. ICOH, PLI, WPLI and LPC all showed a significantly lower match between simulated and empirical FC (around r=0.18) compared to coherence and PLV (Fig. 2.8). For all six metrics, the global correlation was essentially abolished if the underlying SC was shuffled prior to simulation (yellow bars in (Fig. 2.8)). Also, the overall model performance for ICOH, PLI, WPLI and LPC was considerably smaller than the mere correlation between SC and empirical FC (middle row in Fig. 2.5A). What are the possible reasons for this performance drop with ICOH, PLI, WPLI and LPC? One reason could lie in the fact that the reference model SAR does not include delays, thus the simulated FC mainly consists of instantaneous interactions and a comparison with an empirical FC in which those interactions have largely been removed would be futile. However, the results were very similar using the Kuramoto model. The large-scale connectomes derived from all of the four biased metrics did not much reflect the coupling that emerged from our model of fast dynamics based on structural connectivity. Presumably, a considerable amount of functionally relevant synchrony takes place with near zero or zero-phase lag which

is not detected using the biased scores. In fact, zero-phase lag synchronization has been detected between cortical regions in a visuomotor integration task in cats (Roelfsema et al., 1997). More recently, a study of spike train recordings showed how paths among somatosensory areas were dominated by instantaneous interactions (Campo et al., 2015). But synchrony across areas incorporating delays can also lead to high coherence (Gregoriou et al., 2009). A recent modeling study investigated the detection rates of synchrony by different EEG phase synchronization measures (PLV, ICOH, WPLI) in a network of neural mass models. They found that no single phase synchronization measure performed substantially better than all the others, and PLV was the only metric able to detect phase interactions near ±0° or ±180° (Vindiola et al., 2014). This study challenged the supposed superiority of biased metrics in practical applications, because they are biased against zero-phase interactions that do truly occur in the brain. Taken together we argue that by using biased metrics to detect neural synchrony a major portion of relevant coupling is neglected. However, as the relevant stage for comparisons is the source space, the undesired influence of volume conduction effects on the estimated connectivity is partly reduced (Schoffelen and Gross, 2009). Since effects of field spread can never be completely abolished also in the source space, we cannot rule out that volume conduction artifacts have influenced the high correlation in our model. The empirical functional connectome was constructed based on band-pass filtered EEG in the alpha frequency range. Since different FC maps have been detected for different frequency bands (Hipp et al., 2012), it is conceivable that biased vs. unbiased FC metrics might vary in their performance depending on the frequency.

Conclusion

In summary, our framework demonstrates how technical alternatives and choices along the modeling path impact on the performance of a structurally informed computational model of global functional connectivity. We show that determining the resting-state alpha rhythm functional connectome, the anatomical skeleton has a major influence and that simulations of global network characteristics can further close the gap between brain network structure and function.

2.2 Chapter Summary

The manuscript showed that it is possible to bridge the gap from structural data obtained from DTI to functional connectivity data obtained from EEG. Most interesting for the overall topic of this thesis, is the result that the dynamical model of coupled Kuramoto oscillators can explain 54% of the variance in the EEG functional connectivity, while a more simple model of only stationary dynamics can only explain 45% of the variance. This suggests, that dynamic interactions through conduction delays in the network provide additional information to explain the functional EEG data. In comparison, Messé and colleagues found that the SAR model is a better predictor of slow dynamics (< 1 Hz) in functional connectivity obtained from fMRI (Messé et al., 2014). Therefore, our results indicate that fast oscillatory interactions reveal cortical mechanisms that are not covered by correlations in the slowly changing BOLD signal.

Along the processing pipeline from structural data obtained with DTI and functional data obtained with EEG, we evaluated several alternative methods in respect to their ability to close this gap between the structural and functional brain imaging techniques. On the side of the structural data, we showed that additional weighting of the homotopic connectivity to correct for their underestimation (Li et al., 2012a) in fiber tractography can improve the model performance (Finger et al., 2016). On the functional side, we showed that our results are robust for different source reconstruction methods. More importantly, the evaluation of different functional connectivity metrics revealed that coherence and phase-locking values had a much better correspondence between the modeled and empirically measured functional connectivity than metrics excluding zero-lag-interactions. This indicates that zero-lag synchronization plays a crucial role in the dynamics of cortical oscillations. Furthermore, we show that a forward projection of simulated functional activation into sensor space followed by a comparison with EEG functional connectivity in sensor space is highly influenced by mixing of source activation through the lead-field. Therefore, we recommend that the comparison of simulated and empirical functional connectivity should be performed in the source space.

This chapter focused on the modeling of functional connectivity from fixed anatomical structural data. Here, we used empirical measurements of both

structure and function to validate the model performance in predicting the measured functional connectivity based on the measured structural connectivity. In contrast, the next chapter will analyze how the functional activation elicits changes in the anatomical structure as well as in the dynamic interactions on fast timescales and their functional role for binding. There, I will use a bottom-up driven approach by learning the network structure based on activations induced by natural visual stimuli.

Chapter 3

Unsupervised Learning and Temporal Coding

3.1 Phase Synchrony Facilitates Binding and Segmentation of Natural Images in a Coupled Neural Oscillator Network

This section was published as a peer-reviewed article: Holger Finger and Peter König (2013). "Phase synchrony facilitates binding and segmentation of natural images in a coupled neural oscillator network". In: *Frontiers in computational neuroscience* 7.195

Abstract

Synchronization has been suggested as a mechanism of binding distributed feature representations facilitating segmentation of visual stimuli. Here we investigate this concept based on unsupervised learning using natural visual stimuli. We simulate dual-variable neural oscillators with separate activation and phase variables. The binding of a set of neurons is coded by synchronized phase variables. The network of tangential synchronizing connections learned from the induced activations exhibits small-world properties and allows binding even over larger distances. We evaluate the resulting dynamic phase maps using segmentation masks labeled by human experts. Our simulation results show a continuously increasing phase synchrony between neurons within the

labeled segmentation masks. The evaluation of the network dynamics shows that the synchrony between network nodes establishes a relational coding of the natural image inputs. This demonstrates that the concept of binding by synchrony is applicable in the context of unsupervised learning using natural visual stimuli.

Introduction

One of the central questions in neuroscience is how information about a given stimulus is processed in a distributed network of neurons such that it is perceived not only as a collection of unrelated features but as a unified single object. The concept of binding by synchrony has been proposed as a mechanism to coordinate the spatially distributed information processing in the cortex (Milner, 1974; Malsburg, 1981). Experiments in cat visual cortex have confirmed that inter-columnar synchronization indeed corresponds to a relational code that reflects global stimulus attributes (Gray et al., 1989; Singer, 1999; Engel and Singer, 2001). However, the physiological recordings in these early studies were based on the presentation of artificially designed stimuli. In a more recent study Onat, Jancke, and König, 2013 showed in experiments that long-range interactions in the visual cortex are compatible with Gestalt laws. This suggests that the concept of binding by synchrony is also feasible in the case of natural visual stimuli. It is still the center of a heated debate to what extend synchronized activity represents a neural code of binding and segmentation. Esspecially, how the neural system can learn this relational coding when it is exposed to new stimuli is still an open question. The most prominent possibility is that tangential cortico-cortical connections in the visual cortex lead to synchronized activity that implements Gestalt laws. Löwel and Singer, 1992 showed in cats with artificially induced strabismus that selective stabilization of tangential connections occurs between cells that exhibit correlated activity induced by visual experience. Furthermore König et al., 1993 found that the synchronization of cortical activity is impaired in these cats with artificial strabismus. These findings indicate that there is an important interplay between unsupervised learning of tangential connections on behavioral time scales and their role in synchronization phenomena on fast time scales.

The physiological experiments on binding by synchrony have been accompanied by theoretical studies early on. Sompolinsky, Golomb, and Kleinfeld, 1990 investigated how a model of coupled neural oscillators is able to process global stimulus properties in synchronization patterns using abstractly defined neuronal activation levels and predefined coupling strengths for the simulated network. These simulation results showed that the coupling of neural oscillators provides a viable mechanism implementing a coding of perceptual grouping. Such previous work includes studies ranging from networks build out of very simple elements to detailed simulations containing many compartments per unit.

To investigate the functional role of synchronization and its relation to coding, it is important to choose the right level of abstraction in the model. A simplification from detailed spiking neuron models to coupled phase oscillator models allows us to analyze neuronal synchronization in a broader context of a normative model involving unsupervised learning from natural stimuli. A review of these coupled neural oscillator models was done by Sturm and König, 2001, where the authors show the derivation of simplified phase update equations from biologically measurable phase response curves. The simplifications in coupled phase oscillators are based on the assumption that neurons are close to their oscillatory limit-cycle and that a change in the phase of the neuronal inputs induces only a small perturbation to the neuronal phase. The phase update equation in our model is based on the Kuramoto model of coupled phase oscillators (Kuramoto, 1984) in the sense that our model also assumes a very simple sinusoidal phase interaction function. This approximation of the phase interaction by a sinusoidal function allows us to use mathematical simplifications in the simulation of the model.

Very similar to the work of Sompolinsky, Golomb, and Kleinfeld, 1990, we extend the standard formulation of the Kuramoto model with a second variable per neuron to encode the activation of the oscillators. Therefore, in our model the state of a neuron is represented by 2 degrees of freedom, which are separated into activation and phase variables. This discrimination between coding of receptive field features by activation and coding of relationships by phase is a biologically motivated segregation of their different functional roles. Maye and Werning, 2007 specifically compare the synchronization properties of these coupled phase oscillator models with mean-field oscillator models based on the Wilson–Cowan model (Wilson and Cowan, 1972). They state that

the simplified coupled phase oscillators allow decoupling the simulation time constants of fast oscillatory time scales from slow rate coding time scales. Another advantage is an easier analysis of the synchronization patterns, because the direct encoding of the phase variables means that all contextual relationships are coded at the same time. Consequently, we use the dual variable phase model, because it is suitable to answer fundamental questions about the interactions between synchronization phenomena and contextual coding in neural systems.

In contrast to these phase oscillator models, most recent work on segmentation in networks of coupled neural oscillators is based on the so called "local excitatory global inhibitory oscillator network" (LEGION) or similar variants of this model, which was first proposed by Wang and Terman, 1997. In LEGION the dynamics of each oscillatory period of individual units is simulated in detail by time-varying variables describing the internal states of each neuron. In contrast, in our model the oscillatory period is not simulated, but represented only implicitly in the phase variables. Nonetheless, several aspects which we analyze in this work were previously also investiged in LEGION. Namely, similar to Li and Li, 2011 we use a small-world topology, to reduce the computational cost while still allowing binding by synchrony over large distances. We also use parallel computations to speed up the simulations, which was also previously done in LEGION by Bauer et al., 2012.

The above-mentioned previous theoretical studies mostly investigated the processing of artificial stimuli in close analogy to the physiological experiments. These stimuli are heavily dominated by artificial geometric patterns as bars and gratings. However the concept of binding by synchrony makes much more general claims about grouping of sensory representations of natural stimuli. By now a fair number of databases with images considered to be natural is available. However, a problem with generic natural stimuli is that segmentation is not only difficult, but no general ground truth is available. The LabelMe database (Russell et al., 2008) is rather unique, as it contains a large collection of images together with human labeled annotations of image segments. In theoretical studies these labels may serve as a ground truth to evaluate how the relative phases between neurons are coding relational structures on natural stimuli.

The processing of natural stimuli in neural systems can be described as a normative approach in which the representation of the input is learned by an optimization of computational principles (Einhäuser and König, 2010). It has been successfully employed in modeling receptive field properties of simple and complex cells in primary visual cortex. Furthermore, response properties of neurons in higher areas and other modalities have been suggested to follow similar rules. This approach might be extended to include the computational principles that underlie tangential interactions that directly influence synchronization phenomena. This might answer the question whether the concept of binding by synchrony can work in principle with unsupervised learning and natural stimuli.

In this study we investigate whether the concept of binding by synchrony, as has been investigated using abstract stimuli, is viable for natural stimuli. The most important novelty of our approach is the combination of these different concepts described above into one single simulation model to allow the investigation of their interplay: Specifically, we combine normative model approaches of unsupervised learning from natural stimuli with the concept of binding by synchrony in a network of coupled phase oscillators. Importantly, the data driven approach, that utilizes general principles, minimizes the number of heuristics and free parameters. We present large-scale simulations of neural networks encoding real-world image scenes. In the first stage of our algorithm forward projections generate activation levels of neurons corresponding to the primary visual cortex. In the second stage these activation levels are used in a simulation of tangential coupled phase oscillators. We present results with forward projections based on designed Gabor filters that are a good approximation of receptive fields in the primary visual cortex. To allow later canonical generalization in higher network layers, we also present results with forward projections learned in a normative model approach with a sparse autoencoder using natural image statistics. In addition to these learned forward weights, the structural connectivity of the phase simulations is also learned unsupervised using the correlated activity induced by natural stimuli. Performance of the network is tested using images taken from the LabelMe database. Thereby we can investigate how synchronization phenomena might be utilized in sensory cortical areas to bind different attributes of the same stimulus and how it might be exploited for scene segmentation.

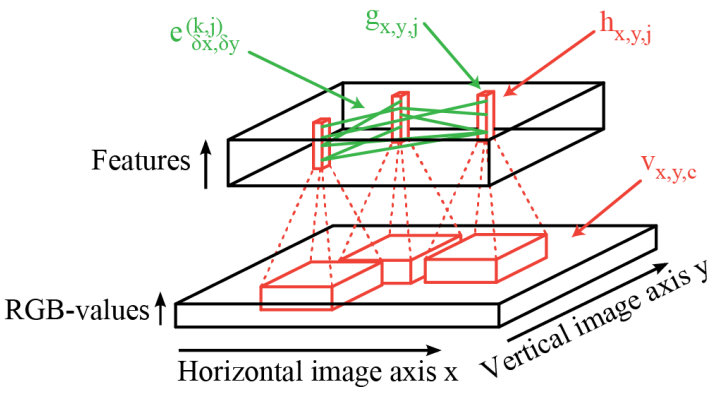


FIGURE 3.1: **Network structure:** Feedforward convolutional filters (red) are applied on the input image $v_{x,y,c}$ (bottom) to generate the activation levels $h_{x,y,j}$ of feature columns (red blocks in the top). These activations $h_{x,y,j}$ are then transformed to activations of oscillators $g_{x,y,j}$ using simple local regularization steps. The intralayer connections $e^{k,j}_{\delta x,\delta y}$ (green) simulate these coupled phase oscillators which synchronize or desynchronize image features.

Material & Methods

The overall network architecture of our simulation model consists of two main parts: 1) Feedforward convolutional filters (red lines in **Figure 3.1**) are used to generate the activation levels for neurons in a layer corresponding to the primary visual cortex. On top of each pixel is a column of neurons which encode different features of a local patch in the input image (black bottom cuboid in **Figure 3.1**). Each feature type is described by a weight matrix which is applied using a 2-dimensional-convolutional operation on each rgb-color-channel of the input image. 2) The obtained activation levels in this 3-dimensional structure (black top cuboid in **Figure 3.1**) are subsequently used to simulate sparse connections (green lines in **Figure 3.1**) between coupled phase oscillators.

We start with the description of the stimulus material (section 3.1). This

is followed by the description of the coupled phase oscillator model (section 3.1) and the sampling mechanism generating the horizontal sparse connections (section 3.1). Afterwards we describe the underlying mechanism of the feedforward generation of activation levels (section 3.1).

Natural Stimulus Material

As stimulus material in our simulations we use images of suburban scenes from the LabelMe database (Russell et al., 2008). Due to computational time constrains we have to restrict the evaluations to a small subset of all available images in the database. In addition, the database is not fixed but new images and segmentation masks are often added. We use only the first 50 images in the folder *05june05_static_street_boston* so that we have a consistent and fixed dataset of well defined images.

These images have initially a resolution of 2560x1920 pixels. We first resize the images to 400x300 pixels to further reduce the computation time of the simulations. Subsequently we subtract the mean pixel values and apply a smoothed zero-phase (ZCA) whitening transformation (Bell and Sejnowski, 1997). For an input image X the whitened pixel values are given by $X_{ZCA} = UDU^T X$, where U is a matrix containing the eigenvectors of the covariance matrix of the image statistics and D is a diagonal matrix with diagonal elements $\frac{1}{\sqrt{\lambda_i + 0.1}}$ where λ_i are the corresponding eigenvalues. This transformation applies local center-surround whitening filters that decrease the correlations in the input images. We implement this whitening transformation using a convolutional image filter.

The images in the LabelMe database come along with human labeled segmentation masks. These segmentation masks correspond to objects that are perceived as a unique concept with an associated abstract label like “tree”, “car” or “house”. We use these supervised segmentation masks for later evaluations of binding in the simulated phase maps. Please note that in our network simulations this segmentation information is not used at any moment in time. Instead, the network connectivity is based solely on unsupervised learning using the statistics of neuronal activations.

Coupled Phase Oscillator Model

Our network of coupled phase oscillators is based on the oscillator model described by Sompolinsky, Golomb, and Kleinfeld, 1990. In the following, we use the same motivational derivation of the phase update equations. We model the probability of firing $P_{x,y,k}(t)$ per unit time of a neuron at image position (x, y) encoding feature type k at time t by an isochronous oscillator. In our simulations we represent the state of the neuronal oscillators by seperated activation variables $g_{x,y,k}$ and phase variables $\Phi_{x,y,k}$. These two variables are linked to the biological interpretation of firing probability by the equation

$$P_{x,y,k}(t) = g_{x,y,k}(1 + \lambda \cdot \cos(\Phi_{x,y,k}(t))), \tag{3.1}$$

where the parameter $0 < \lambda < 1$ controls the relative strength of the temporal oscillation in relation to the overall firing probability of the neuron. The phase progression is a periodic function $\Phi_{x,y,k}(t) = \Phi_{x,y,k}(t + 2\pi)$. In our work, the calculation of the activation levels $g_{x,y,k}$ significantly differs from the simple artificial tuning curves used in Sompolinsky, Golomb, and Kleinfeld, 1990. A detailed description of how these activation levels are obtained will be presented in section 3.1. The activation levels $g_{x,y,k}$ are normalized by dividing by the local sum of all activation levels at each image position such that $\sum_k g_{x,y,k} = 1 \ \forall x, y \in \mathbb{Z}$. In the simulations presented in this work the activation levels of each neuron are only computed once from the input image using feedforward projections (red lines in **Figure 3.1**) and are then kept constant during the simulation of the phase model. This simplification of constant activation levels is based on the assumption that the stimulus presentaion on behavioral timescales ($\approx$ seconds) remains constant during the phase synchronization which happens at very fast timescales (i.e. gamma frequency $\approx$ 40 Hz). Another argument to support this assumption is that the visual cortex seems to operate in a regime of self-sustained activity (Stimberg et al., 2009) and therefore we can assume constant activation levels during the phase simulation.

After these activation levels V are computed, we simulate the horizontal coupling between the phase oscillators. The phase connections in our network are described by a weighted graph $G = (H, E)$ where the neurons $g_{x,y,j} \in H$ are the vertices organized in a three dimensional block (**Figure 3.1**). An edge

$e^{(j,k)}_{\delta x,\delta y} \in E$ describes synchronizing (positive) or desynchronizing (negative) connections from neurons $g_{x,y,j}$ to neurons $g_{x+\delta x,y+\delta y,k}$. The phase of each neuron is then modeled according to a differential equation describing weakly coupled phase oscillators (Kuramoto, 1984)

$$\frac{d\Phi_{x,y,k}(t)}{dt} = \omega - \frac{1}{\tau} \sum_{e^{(j,k)}_{\delta x,\delta y} \in E} \Big(g_{x,y,k} \cdot e^{(j,k)}_{\delta x,\delta y} \cdot g_{x-\delta x,y-\delta y,j} \cdot \sin(\Phi_{x,y,k}(t) - \Phi_{x-\delta x,y-\delta y,j}(t)) \Big), \tag{3.2}$$

where τ is the time scale of the phase interactions and ω is the natural frequency of the modeled neural oscillations. We assume that all neurons have the same intrinsic natural frequency ω and the interaction strength $g_{x,y,k} \cdot e^{(j,k)}_{\delta x,\delta y}$. $g_{x-\delta x,y-\delta y,j}$ is proportional to the activation levels of the pre- and postsysnaptic neurons. Note that our model is in contrast to the more common formulation of the Kuramoto model with heterogeneous frequencies and fixed homogenous all-to-all interaction strengths.

A major difference to the phase update equation used in Sompolinsky, Golomb, and Kleinfeld, 1990 is that we neglect the noise term in the differential equation of each oscillator. The noise term in Sompolinsky, Golomb, and Kleinfeld, 1990 is used as the primary source of desynchronization in the network. In contrast, in our work, we use a normative model to learn not only synchronizing but also desynchronizing connections (see section 3.1). For an easier analysis and interpretation of the results, it is advantagous to have only a single source for the desynchronization in the network. Therefore we decided to use a deterministic phase model, although it was previously shown that noise is an important factor to control the network coherence. In addition to a simpler interpretation it reduces the number of model parameters and is also more compatible to further applications of gradient descent learning to change the strength of the phase interactions.

We can further simplify the equation by using the fact that we model isochronous oscillators with homogeneous frequencies. In equation 3.2 all phase variables $\Phi_{x,y,k}(t)$ have a constant phase progression with frequency ω. We can use a simple transformation to a new variable, which represents only the phase

offsets between neurons:

$$\varphi_{x,y,k}(t) = \Phi_{x,y,k}(t) - \omega t. \tag{3.3}$$

This new phase variable $\varphi_{x,y,k}(t)$ describes the relative phase of neuron k to the global fixed network oscillation with frequency ω. Substitution into the equation above leads to a simplified phase update equation

$$\begin{aligned} \frac{d\varphi_{x,y,k}(t)}{dt} = & -\frac{1}{\tau} \sum_{e^{(j,k)}_{\delta x,\delta y} \in E} g_{x,y,k} \cdot e^{(j,k)}_{\delta x,\delta y} \cdot g_{x-\delta x,y-\delta y,j} \\ & \cdot \sin(\varphi_{x,y,k}(t) - \varphi_{x-\delta x,y-\delta y,j}(t)). \end{aligned} \tag{3.4}$$

In this equation it can be seen that the timescale τ of the phase interaction strength is decoupled from the oscillatory timescale $1/\omega$. Please also note, that a change of the parameter τ would not qualitatively change the results of our simulations. Instead it would just linearly change the units of the time axes. Therefore we show the simulation results with the time axis measured in iterations, which could be linearly scaled to arbitrary time units to best fit to different biological measurements.

This phase update equation is used in our simulations to model the horizontal connections in the network. It allows directly specifying synchronizing interactions from neuron $g_{x,y,j}$ to neuron $g_{x+\delta x,y+\delta y,k}$ with a positive connection weight $e^{(j,k)}_{\delta x,\delta y}$ and desynchronizing interactions with a negative weight respectively. We simulate these coupled differential equations using a 4th-order Runge-Kutta method.

Horizontal Interaction Strengths

We use correlation statistics of the induced activation levels to set the intralayer connection strengths similar to a simple Hebbian learning rule. We write $\rho^{(k,m)}_{x,y}$ to denote the Pearson cross-correlation between the activations of feature type k at image position $(\tilde{x}, \tilde{y})$ and the activations of feature type m at the shifted image position $(\tilde{x}+x, \tilde{y}+y)$. Each correlation value in this tensor is calculated from the correlation statistics over approximately 1 million network activations induced by 50 natural images and presented at 236x86 image positions.

These horizontal connections make up the coupling between the neural oscillators. Instead of full connectivity, we use stochastically sampled sparse directed connections from the correlation matrix. To exclude noise in the correlation matrix, we use the Benjamini–Hochberg–Yekutieli procedure (Benjamini and Yekutieli, 2001) under arbitrary dependence assumptions with a false-discovery rate of 0.05.

The probability of a positive (+1) or a negative connection (-1) in the connectivity graph $G = (H, E)$ is then given by

$$P\left(e_{x,y}^{(j,k)} = \pm 1\right) = \eta_{\pm} \cdot \frac{\max\left(0, \pm\rho_{x,y}^{(j,k)}\right)}{\sum_{\tilde{x},\tilde{y},m} \max\left(0, \pm\rho_{\tilde{x},\tilde{y}}^{(m,k)}\right)}, \tag{3.5}$$

where η_+ specifies the total number of afferent synchronizing connections and η_- the total number of afferent desynchronizing connections per neuron. Therefore synchronizing connections exist only between naturally correlated features and desynchronizing connections between anti-correlated features.

We sample this sparse tangential connection pattern such that it is invariant to spatial shift transformations. The convolutional structure of the forward projections leads to activation and phase variables that are stored in a 3-dimensional block (top of **Figure 3.1**) with two dimensions given by the spatial extend of the image and one feature dimension. This convolutional structure can be exploited for the sparse horizontal connections to significantly speed up the computation. Therefore we specify the properties of the coupled oscillator connections only for a generic feature column. These connections are then applied at each image position. Specifically, in our implementation each sampled tangential connection is specified by 5 variables: the horizontal and vertical connection length in image directions and the indices of the afferent and efferent feature maps and the connection weight. This has the advantage that the phase update equation can be implemented as a vectorized convolutional operation although the connection pattern is highly sparse.

Feedforward Connectivity

We compare the binding and segmentation performance of the coupled neural oscillator model using two different ways to generate the activation levels for the neurons. We first describe hand-crafted feedforward Gabor weights (section 3.1) and then the unsupervised learning of receptive fields using a convolutional autoencoder (section 3.1). Finally, activation functions are presented to further regularize the resulting feature representations (section 3.1).

Gabor Filters For reference we use a set of Gabor filters with specified orientation, frequency and color tuning to generate the activation levels for the phase simulation. Thereby we can analyze the phase oscillator network based on a regularly defined set of features that can be parameterized.

We generate linear convolutional weights (marked in red in **Figure 3.1**) using an approximate Gaussian derivative model, which was shown to be a good fit for the receptive fields of simple cells in the primate visual cortex (Young, 1987). We use only non-directional three-lobe monophasic receptive fields (Young and Lesperance, 2001) to reduce our model parameters. We implement the Gaussian derivative model using difference-of-offset-Gaussians with a slightly larger center compared to surround to code color offsets. The receptive fields that are used in our simulations have a size of 12x12 pixels and are defined by

$$W_{x,y} = g_{2\sigma}(y) \cdot \Big(-5 \cdot g_{\sigma}(x+\sigma) + 10.1 \cdot g_{\sigma}(x) - 5 \cdot g_{\sigma}(x-\sigma)\Big), \quad (3.6)$$

where $g_{\sigma}(x)$ is a one dimensional Gaussian distribution with standard deviations $\sigma = 1.5$ pixels (or $g_{2\sigma}(y)$ with standard deviation of $2\sigma = 3$ pixels) and the coordinates x and y are rotated giving a total of 8 orientations in steps of $22.5°$. The convolutional filters are applied to the images with a stride of 2 pixels in both image dimensions and are followed by a sigmoidal activation function to scale the values to a reasonable interval between 0 and 1. We apply each orientation filter separately to all color channels (red, green, blue). Furthermore we add features for the complementary color channels similar to the on-off discrimination in the visual pathway from the retina to the visual cortex. The direct linear dependency between these pairs of opponent-color channels is removed later with additional activation functions described in section 3.1.

In summary, we have a total of 48 convolutional feature channels per image position: 8x orientations, 3x rgb-color channels, 2x opponent-color channels. This overcomplete neural representation of the input images is used to generate the activation levels for the phase simulations.

Cortical measurements show that the distribution of non-directional monophasic simple cells is roughly uniformly distributed between zero-, first- and second order Gaussian derivatives (Young and Lesperance, 2001). We performed the simulations presented here also with mixed receptive fields of zero-, first- and second-order Gaussian derivatives and obtained similar results. We present here only results with second order Gaussian derivatives, because this reduces the number of model parameters drastically.

Autoencoder Filters As a comparison to these regular hand-designed Gabor filters we analyze the oscillatory network based on activation levels generated by unsupervised learned autoencoder weights. A good overview of the concepts described in this section can be found in Ngiam et al., 2011, where the authors analyze different optimization methods for convolutional and sparse autoencoders. An autoencoder learns a higher level representation from the stimulus statistics such that the input stimuli can be reconstructed from the hidden representations. In additon, we optimize the sparsity of the activation levels in this representation, which was shown to learn connection weights which resemble receptive fields in the visual cortex (Olshausen, 1996; Hinton, 2010; Le et al., 2011).

A common trick in unsupervised learning in neural networks are shared connection weights to reduce the number of parameters that have to be learned, which can be accomplished by a convolutional feed-forward network in the case of images (LeCun et al., 1998; Lee et al., 2009). The structure of our convolutional autoencoder is shown in **Figure 3.2**. The feedforward projections that generate the activation of feature map j consist of convolutional filters $W_{x,y,c,j}$ (red lines in **Figure 3.2**) with input features $c \in \{1, 2, 3\}$ (rgb-colors) and a bias term b_j and is followed by a sigmoidal activation function. Therefore the hidden layer activation map of feature $j \in \{1, 2, ..., J\}$ is described by

$$h_{x,y,j} = f\Big(\sum_{c=1}^{3} W_{x,y,c,j} * v_{x,y,c} + b_j\Big). \tag{3.7}$$

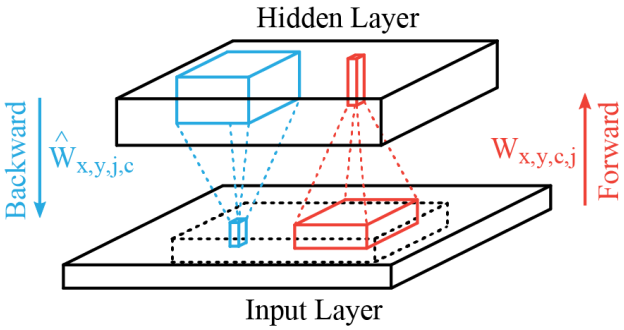


FIGURE 3.2: **Structure of the convolutional autoencoder:** Convolutional forward weights (red) compute the hidden layer activation levels and convolutional backward weights (blue) generate the reconstruction of the given input. The reconstruction layer is compared to the centered part (dashed block) of the input layer.

The hidden layer activation h of each input image sample is also a 3 dimensional block (horizontal and vertical image dimensions and the feature type). The weight matrix W is a 4 dimensional structure which describes the connection weights from a convolutional input block to one output column in the hidden layer. The convolutional image operations ($*$) are applied in the image directions x and y between all combinations of input feature maps c and all output feature maps j.

We use linear activation functions for the backward projections (blue lines in **Figure 3.2**) so that the output matches the scale of the input images (zero-mean). We use another set of weights $\hat{W}_{x,y,j,c}$ and bias terms $\hat{b}_c$ to describe these backward connections. Therefore the activation in the reconstruction layer is given by

$$\hat{v}_{x,y,c} = \sum_{j=1}^{J} \hat{W}_{x,y,j,c} * h_{x,y,j} + \hat{b}_c, \tag{3.8}$$

where $J = 100$ is the number of different feature types. During the learning stage only the valid part (no zero padding) of the convolutions are used for the forward and backward projections to avoid edge effects of the image borders

on the learned weights. Similar to the Gabor filters the convolutional filters have a size of 12x12 pixels and are applied using a stride of 2 pixels leading to a reduction in the resolution of the hidden layer.

We use the sum of 3 optimization functions to learn the forward and backward weights of the autoencoder. The first optimization term which is minimized is the reconstruction error averaged over all positions and training samples s and is given by

$$\Psi_1 = \left\langle \frac{1}{2} \left\| \hat{v}^{(s)}_{x,y,c} - v^{(s)}_{x,y,c} \right\|^2 \right\rangle_{x,y,s} . \tag{3.9}$$

The second term optimizes the sparseness of the hidden units as described by Hinton, 2010 and Le et al., 2011 with

$$\Psi_2 = \beta \cdot \sum_j \mathrm{KL}\left(\tilde{h} \middle\| \langle h^{(s)}_{x,y,j} \rangle_{x,y,s} \right), \tag{3.10}$$

where KL is the Kullback-Leibler-divergence between two Bernoulli distributions with expected values $\tilde{h}$ and $\langle h^{(s)}_j \rangle_s$. We set the desired average activation $\tilde{h} = 0.035$.

The third term is a weight decay (L2-norm) of all forward and backward weights and is given by

$$\Psi_3 = \frac{\lambda}{2} \cdot \Big(\sum_{x,y,c,j} W^2_{x,y,c,j} + \sum_{x,y,j,c} \hat{W}^2_{x,y,j,c} \Big). \tag{3.11}$$

This optimization term pushes all connection weights towards zero such that only the connections which help to extract useful features remain. Therefore it provides a regularization mechanism during learning.

For the simulations presented in this paper we use a relative weighting between these optimization functions given by $\beta = 90$ and $\lambda = 0.3$. The gradients of the optimization functions are calculated using back propagation of error signals and were checked using numerical derivatives. The sum of the three terms described above is minimized with the limited memory Broyden-Fletcher-Goldfarb-Shanno algorithm (L-BFGS), which uses an approximation to the inverse Hessian matrix (Liu and Nocedal, 1989). We use the *minFunc*

library of Mark Schmidt[1] with default parameters for line search with a strong Wolfe condition. We use L-BFGS because it converges much faster in comparison to standard gradient descent, especially in the case of autoencoders with sparseness constrains (Ngiam et al., 2011). Another advantage of L-BFGS is that extensive tuning of learning parameters as in standard gradient descent methods is not necessary.

The training data consists of 1000 color patches (60x60 pixels) sampled from the folder *05june05_static_street_boston* of the LabelMe database (Russell et al., 2008). This corresponds to 625.000 training samples per convolutional fragment where the forward weight matrix is applied. After 500 iterations the features are mostly oriented patches and sensitive to different colors.

Regularization of Activation Levels Although the Gabor and autoencoder filters are both followed by a sigmoidal activation function, we further sparsify the activation levels $h_{x,y,k}$ with feature types $k \in \{1..K\}$ in a similar way to local cortical circuitry. We want to constrain the number of active neurons, rather than the mean activation levels. Therefore we subtract at each image position the average local activation levels. Subsequently a half-wave rectification is applied to constrain the activation levels again to the positive domain with roughly half of the neurons inactivated:

$$\tilde{h}_{x,y,k} = \max\left(0, h_{x,y,k} - \sum\nolimits_{j=1}^{K} h_{x,y,j}\right). \tag{3.12}$$

Consequently the hard sparseness (Rehn and Sommer, 2007) is artificially increased and these inactivated neurons do not take part in the coupling of phase oscillations (see section 3.1). Thereby the number of possible interactions in the phase simulations is reduced.

As a last step we have to normalize the activation levels at every image position similar to local contrast adaptation in the visual system. We want to make sure that the overall local activation is uniform over the visual field such that an efficient coding of regions of high contrast and regions of low contrast is possible simultaneously. Therefore we divide all activation levels by the sum

[1] `http://www.di.ens.fr/~mschmidt/Software/minFunc.html`

of activations over all features at each image location:

$$g_{x,y,k} = \frac{\tilde{h}_{x,y,k}}{\sum_{j=1}^{K} \tilde{h}_{x,y,j}}. \tag{3.13}$$

As a result we have sparse activation maps with a large proportion of inactive neurons and the same average local activations at all image positions.

Results

In a first step we analyse the properties of the activation patterns induced by the natural images (section 3.1). Subsequently we evaluate the correlation statistics of these induced feature activations (section 3.1) and the resulting sparse connectivity pattern (section 3.1). Based on this connectivity pattern we show simulations of the coupled phase oscillator model and the resulting dynamic phase maps (section 3.1). Finally, evaluations of these binding maps are presented based on human labeled segmentation masks (section 3.1).

Sparseness of Activation

The simulation of the coupled phase oscillators is based on the activation levels that were generated from natural images. The phase coupling is highly dependent on the type of feature representation that is used to generate the activation levels. The first reason is that the connectivity is based on the correlation between features. The second reason is that also the actual strength of the dynamic coupling is proportional to the current activation levels. Therefore the statistics of activation plays a crucial role in the formation of the dynamic binding maps.

Hand labeled photographs of suburban scenes from the LabelMe database (Russell et al., 2008) are used to generate feature representations with the linear convolutional forward weights followed by a sigmoidal function. The linear convolutional kernels of the Gabor receptive fields contain only one spatial frequency and equally spaced orientations (**Figure 3.3A**). In contrast, the learned weights of the sparse autoencoder (**Figure 3.3B**) cover a diverse set of spatial frequencies, colors and orientations.

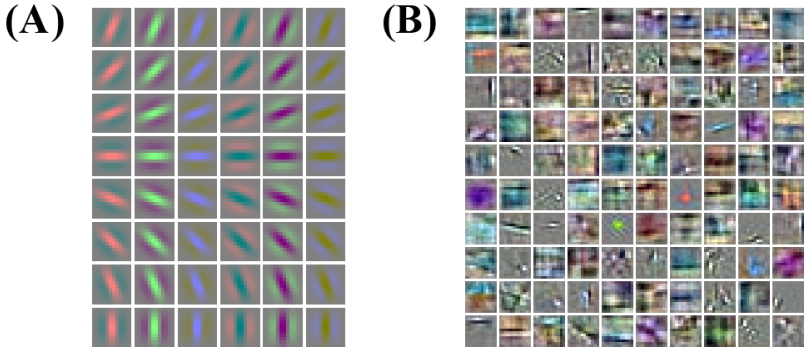


FIGURE 3.3: **Receptive fields of the feed-forward connections generating the activation levels for the phase simulations.** (A) The regular Gabor filters are generated with 8 different orientations and 6 different color channels. (B) The convolutional autoencoder weights are learned by optimizing the reconstruction cost, sparseness and weight decay.

We compare the activation levels of features obtained with the regular Gabor weights and the autoencoder weights. A very important characteristic of neuronal activations is the level of sparseness. A high level of activation sparseness means that the neuron is most of the time very silent and only rarely very active. This analysis of sparse coding should not be confused with the graph theoretic sparseness which will be analysed in section 3.1. A qualitative comparison of the activation histograms (**Figure 3.4A**) shows that the autoencoder activations are sparser compared to the Gabor activations. The phase model is based on the assumption that the activation is restricted to the positive domain. Note that this is in contrast to many normative models of early visual processing which assume a feature code with a Gaussian distribution with zero mean. Furthermore, in our model we are mostly interested in the "hard sparseness" of the activation levels, meaning that the activation is most of the time exactly zero and only rarely very high (Rehn and Sommer, 2007). A comparison with a Gaussian distribution restricted to the positive domain with the same mean (dashed line in **Figure 3.4A**) reveals that the feature activations after the sigmoidal activation function are not necessarily sparse in the context of a positive distribution with this hard sparseness criteria.

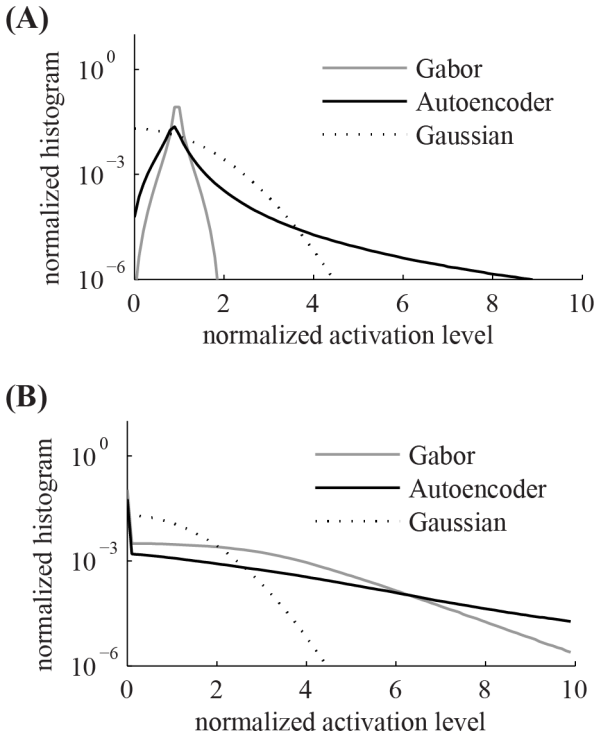


FIGURE 3.4: **Histogram of activation levels averaged over all feature types.** The distributions of activation levels are compared to a Gabor distribution. (A) After sigmoidal activation function. (B) After mean subtraction, half-wave rectification and division by the sum.

The sigmoidal activation function is followed by the subtraction of mean, rectification and the division by the sum over all features. The resulting histograms of these activation levels (**Figure 3.4B**) show an increased hard sparseness for both types of receptive fields. These additional preprocessing steps are similar to local regulatory mechanisms in the cortex.

A quantitative evaluation of the sparseness of the activation levels is given by the kurtosis. We use the standard measure of excess kurtosis but without mean normalization because the phase model assumes a non-negative feature coding by activation. Therefore we evaluate the hard sparseness of feature type j with activation levels $h^{(s)}_{x,y,j}$ by the kurtosis of a zero-centered distribution given by

$$\text{kurt}_j = \frac{\left\langle \left(h^{(s)}_{x,y,j} \right)^4 \right\rangle_{x,y,s}}{\left(\left\langle \left(h^{(s)}_{x,y,j} \right)^2 \right\rangle_{x,y,s} \right)^2} - 3, \tag{3.14}$$

where $\langle . \rangle$ is the mean over all image positions (x, y) and image samples s from the labelMe database. The estimated median kurtosis over all receptive field types increases for the activations g after the normalization steps described above in comparison to the activations h before the normalizations (**Table 3.1**). A comparison with a Gaussian distribution, which has a kurtosis of 0, reveals that the additional activation functions indeed increase the sparseness and lead to a leptokurtic distribution of activations. Overall the activations generated by the autoencoder are more sparse in comparison with the hand designed Gabor filters.

The additional activation functions are crucial for the subsequent phase simulations. The mean subtraction and half-wave rectification increase the hard sparseness of activations. This reduction in the number of active neurons leads to a reduction in the number of active tangential phase connections. Therefore the features in the input image do not only multiplicatively modulate the strength of the phase interaction but also deactivate many phase connections entirely leading to a completely new effective tangential connectivity pattern.

TABLE 3.1: **Median kurtosis of feature activations**

	After sigmoid Activations $h_{x,y,k}$	After normalizations[1] Activations $g_{x,y,k}$
Gabor	-1.96	2.61
Autoencoder	-0.63	11.62

[1]After the subtraction of mean, half-wave rectification and division by the local sum of the new activation levels.

Statistics of Horizontal Cross-Correlations

The horizontal connections between the coupled phase oscillators are sampled from the cross-correlations of induced activation levels as described in equation 3.5. Therefore we describe the horizontal correlations in this section and evaluate the anisotropy of receptive field types. The 4 dimensional cross-correlation tensors $\rho_{x,y}^{(k,m)}$ as defined in section 3.1 are shown in **Figure 3.5** for 8 feature types. The Gabor receptive fields have a more regular correlation matrix (**Figure 3.5A**) compared to the learned autoencoder receptive fields (**Figure 3.5B**). The correlations between the activations of Gabor receptive fields are itself similar to high frequency Gabor functions. In contrast, the receptive fields learned by the autoencoder capture different spatial frequencies and a variety of different colors which is also reflected in the spatial cross-correlations. In both cases the horizontal cross-correlations extend over visual space up to three times the receptive field size. This suggests that the correlations indeed comprise higher-order correlation statistics of the natural images and not only interactions between overlapping receptive fields.

To analyze and compare the correlation tensor of the autoencoder and the Gabor filters, we calculate statistics for different correlation distances in visual space. The indices of the tensor are illustrated in the schematic in **Figure 3.6A**. For each distance r in visual space we calculate statistics over $\rho_j^{(k,m)}$ where

$$j \in R_r := \left\{ (x,y) \in \mathbb{Z}^2 \middle| r - \frac{1}{2} \leq \sqrt{x^2+y^2} < r + \frac{1}{2} \right\}. \tag{3.15}$$

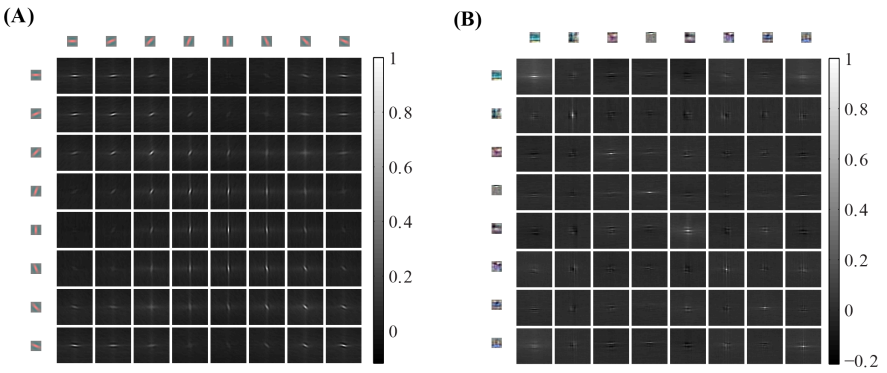


FIGURE 3.5: **Cross-correlations between different feature activations shifted in visual space.** The shown cross-correlations are based on the activation levels induced by natural images. Only a subset of 8 features is shown. The patches on the top and left row show the forward weight matrix of the receptive fields. The other patches show the spatial correlation between these features. The feature weights are shown at the same spatial scale as the shifts in the cross-correlations. (A) The correlations between 8 oriented Gabor filters of one of the 6 color channels are shown. (B) The correlations between 8 randomly choosen autoencoder features are shown.

The mean absolute value of the cross-correlations decreases for larger correlation distances r as shown in **Figure 3.6B**. The mean standard deviation of these absolute correlation values over different spatial directions also decreases but with a steeper slope (**Figure 3.6C**). To make a relative statement about the isotropy in the correlation tensor we also calculate the coefficient of variation over different directions. Therefore we define the average anisotropy at radius r as

$$\text{anisotropy}(r) := \left\langle \frac{\text{std}_{j \in R_r}\left(\rho_j^{(k,m)}\right)}{\text{mean}_{j \in R_r}\left(\rho_j^{(k,m)}\right)} \right\rangle_{k,m} \tag{3.16}$$

This mean anisotropy averaged over all pairs of receptive field types has a local maximum at visual distances of around 8-10 pixels (**Figure 3.6D**). This suggests that the short range phase connections over this distance help more in the synchronization of fine structures. The anisotropy has a local minimum at

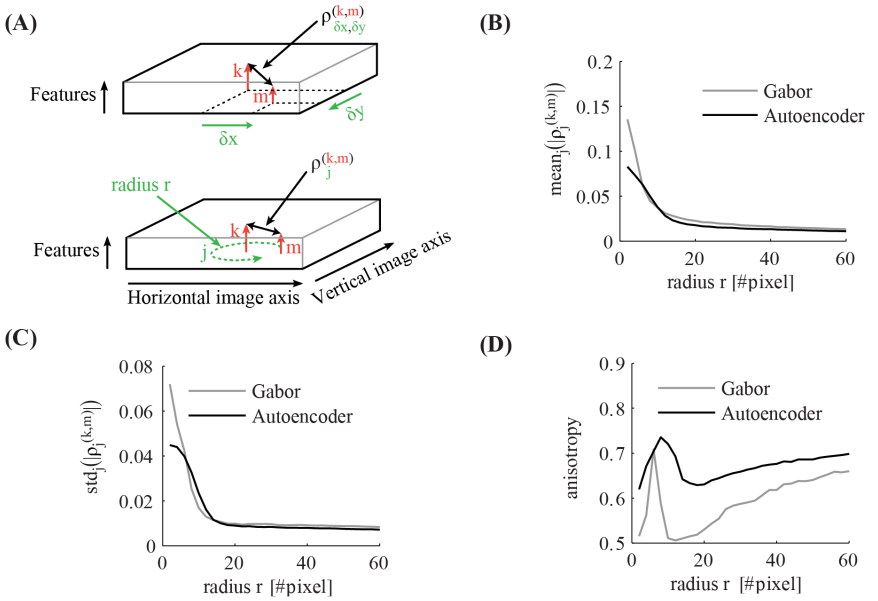


FIGURE 3.6: **The statistics of the correlation matrix evaluated for different distances r in visual space.** (A) Schematic to illustrate the indices of the correlation tensor. In the top schematic the correlation tensor is indexed by horizontal (x) and vertical (y) offsets in visual space. In the bottom schematic the correlation tensor is indexed by $j \in R_r$ for a certain distance r in visual space. The other panels compare the correlation tensor of Gabor filters (gray) and autoencoder filters (black) for different distances r. All shown statistics are averaged over all pairs of receptive field types k and m. (B) Mean over all directions. (C) Standarddeviation over different directions for a certain pair of feature types. (D) The anisotropy averaged over all pairs of receptive fields as described in the main text.

distances around 15-16 pixels, where more long range phase connections are dominantly used to fill-in segment pixels with similar colors.

Sparsely Connected Oscillator Network

The correlation values are used to sample the sparse connections for the simulations of coupled phase oscillators. We restrict the sampled connectivity pattern in simulations of natural scenes to 200 synchronizing and 200 desynchronizing afferent connections per neuron if not stated otherwise. The phase simulations of natural image scenes are run in a network of 200x150x48 neurons for Gabor features or 200x150x100 for autoencoder features respectively. Therefore the percentage of connections that are actually formed compared to all possible connections assuming full connectivity is approximately 0.014 % in the case of Gabor features and 0.007 % for autoencoder features. Thus, this procedure leads to a very sparse connectivity in comparison to a network of all-to-all interactions.

We evaluate the sampled connectivity based on natural image statistics using graph theoretic measures. The connectivity structure is represented as a graph $G = (H, E)$ as described in section 3.1. We compute the statistics not only over the graph of all connections E but also for the subgraph of synchronizing connections $E^+ := \left\{ e_{x,y}^{(j,k)} \in E \middle| e_{x,y}^{(j,k)} = +1 \right\}$ and the subgraph of desynchronizing connections $E^- := \left\{ e_{x,y}^{(j,k)} \in E \middle| e_{x,y}^{(j,k)} = -1 \right\}$ individually.

For a graph with edges E we calculate the fraction of intra-feature connections as

$$\mu = \frac{\left| \left\{ e_{\delta x,\delta y}^{k,m} \in E \middle| k = m \right\} \right|}{\left| \left\{ e_{\delta x,\delta y}^{k,m} \in E \middle| k \neq m \right\} \right|} \cdot 100\%. \tag{3.17}$$

The most obvious observation is that the fraction of intra-feature connections is larger for synchronizing connections in comparison to the desynchronizing connections (**Table 3.2**). The reason is that positive correlations, which are used to sample these synchronizing connections, are stronger between the same feature type shifted over visual space. In contrast negative correlations and thus desynchronizing connections are less likely to occur between the same feature type shifted over visual space. Another observation is that the

fraction of intra-feature connections of the Gabor features is roughly twice as large as in the case of the autoencoder features. The reason is that we use 100 autoencoder features and only 48 Gabor features while the total number of sampled synchronizing and desynchronizing connections per feature remains constant.

A more elaborate evaluation of the sampled connectivity of our network can be done using the clustering coefficient and the small-world characteristics (Watts and Strogatz, 1998; Humphries, Gurney, and Prescott, 2006), which are also shown in **Table 3.2**. To define the local clustering coefficient in an infinite graph $G = (H, E)$, we analyze the connectivity of the neurons in a generic feature column at position $(x, y) = (0, 0)$. We define the neighbors of neuron $g_{0,0,k}$ coding feature type $k \in \{1..K\}$ as the set of all neurons which are directly connected in the graph as

$$N_k = \left\{ g_{x,y,m} \in H \middle| e^{k,m}_{x,y} \in E \vee e^{m,k}_{-x,-y} \in E \right\}, \tag{3.18}$$

where we consider outbound ($e^{k,m}_{x,y}$) and inbound ($e^{m,k}_{-x,-y}$) connections of the neuron. Then we define the local clustering coefficient of a feature type k in our network as the fraction of the number of direct connections between neighbors to the number of pairs of neighbors:

$$\gamma_k = \frac{|\{e^{m,n}_{x,y} \in E | g_{\tilde{x}+x,\tilde{y}+x,m} \in N_k \wedge g_{\tilde{x},\tilde{y},n} \in N_k\}|}{|N_k| \cdot (|N_k| - 1)} \tag{3.19}$$

We show the global clustering coefficients $\gamma =< \gamma_k >_k$ for our sampled networks comprising only the synchronising, only the desynchronizing or all connections in the second row of **Table 3.2**.

The evaluation of the graph comprising all connections shows that the mean clustering coefficient is roughly the same for the Gabor and the autoencoder features. But the evaluation of graphs individually reveals that the clustering coefficient of only the synchronizing graph is higher for the Gabor features in comparison to the autoencoder features. And reciprocally, the desynchronizing connections show a stronger clustering in the case of autoencoder features. An explanation for this difference is that the autoencoder learns a more diverse set of receptive fields by optimizing the reconstruction error. In comparison, the regular Gabor receptive fields cover only predefined colors, spatial frequencies

and orientations, which are not optimized to cover a broad range of statistics in the input images. Therefore the correlation structure in the Gabor activations shows stronger clustering. For comparison, we also show the corresponding clustering coefficients γ_{random} of the equivalent networks with the same connection lengths (measured in pixel distance) but rotated by random angles and connected to random features.

We can further use the small-world index to measure the capability of neurons in our network to reach other neurons via a small number of interaction steps. The small-world index is a quantitative definition of the presence of abundant clustering of connections combined with short average distances between neuronal elements, proposed by Humphries, Gurney, and Prescott, 2006. It can characterize a large number of not fully connected network topologies. The connectivity within the 3-dimensional grid of our model is sampled such that it is invariant to shifts in the two image dimensions. Therefore we have to slightly adapt the small-world index for our infinite horizontal sheet consisting of feature columns with identical connection patterns. We use the definition of the small-world index

$$\sigma_{\text{sw}} = \frac{\gamma / \gamma_{\text{random}}}{\lambda / \lambda_{\text{random}}}, \tag{3.20}$$

where the shortest path lengths λ and λ_{random} measure the number of network hops needed to connect two neurons within our sampled network and a random network respectively. We use the average over all shortest path lengths between all pairs of neurons within one feature column. A network graph must have a small-world index σ_{sw} larger than one to meet the small-world criteria. The evaluations show that the graph comprising the synchronizing connections exhibits small-world properties while the desynchronizing connections are closer to a random connectivity and do not exhibit small-world properties (**Table 3.2**). The small-world property might be helpful in the synchronization of distant neurons.

Phase Simulations

The resulting connectivity pattern is used in the phase simulations. All shown simulations of the coupled phase oscillator networks are initialized with random phase variables. The activation levels are only set once in the beginning

TABLE 3.2: **Graph theoretic statistics of the sparse connectivity pattern**

	Gabor			Autoencoder		
	All E	Sync. E^+	Desync. E^-	All E	Sync. E^+	Desync. E^-
fraction of intra-feature connections μ	6.49 %	12.87 %	0.10 %	3.12 %	6.17 %	0.07 %
global clustering coefficient γ (in 10^{-3})	3.21	3.53	0.64	3.18	2.49	1.27
global clustering coefficient random γ_{random} (in 10^{-3})	2.35	1.11	1.25	2.18	1.03	1.18
mean shortest path length λ	2.01	2.60	2.46	2.17	2.63	2.59
mean shortest path length random λ_{random}	2.02	2.53	2.47	2.15	2.62	2.56
small world index σ_{sw}	1.37	3.09	0.51	1.44	2.39	1.07

and remain the same throughout the phase simulations. During the simulations attractors are formed in the phase space and are localized in certain image regions.

A simulation of the coupled phase oscillator model with localized connectivity and with uniform activation levels shows that pinwheel structures will form in the phase map (**Figure 3.7A-B**). The connectivity length in the network determines the scale of the pinwheels. During the simulation these pinwheels attract each other and annihilate (Wolf and Geisel, 1998). The probability of the formation of pinwheels decreases for network connectivity patterns that are less locally dense but more sparse and spread out.

In the next simulations we use several feature types to encode different aspects of the input images. To visualze the resulting 3-dimensional structure of phase variables $\varphi_{x,y,k}$ we calculate the circular mean at each image position weighted by the corresponding activation levels:

$$\varphi_{x,y}^{\text{avg}} := \arg\left(\sum_k g_{x,y,k} e^{i\varphi_{x,y,k}}\right), \tag{3.21}$$

where arg is the complex argument. We show the average phase variables $\varphi_{x,y}^{\text{avg}}$ coded as color hue to visually represent the circular structure of the phase.

Simulations with simple grayscale objects as inputs show that Gestalt laws of grouping are implemented in the phase network (**Figure 3.7C-D**). Both simulations are based on gabor receptive fields with horizontal connectivity obtained from statistics of natural images. A simulation of two colinear aligned bars shows that the phase variables implement the law of continuity (**Figure 3.7C**), because the two bars synchronize although there is a light patch in between the two bars. A simulation of a dashed black circle shows the law of closure (**Figure 3.7D**), because all segments of the circle are synchronizing to the same phase values. The synchronized state of the circle means that the phase variables at different segments of the circle code the global attribute and bind the individual circle segments together. Depending on the initialization of random phase variables, cases exist where the circle does not synchronize to one coherent phase but forms a continuous phase progression one or multiple times from 0 to 2π. On one hand these simulations reproduce the previous studies demonstrating binding properties of coupled neural oscillators. On the other

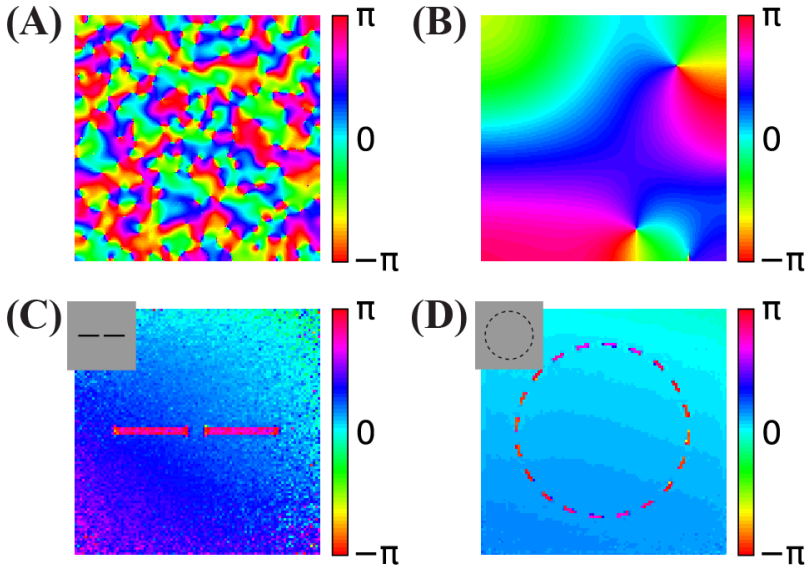


FIGURE 3.7: **Simulations of the coupled phase oscillator model.** The phase variables are shown as color hue. (A,B) Phase simulations of a 2-dimensional grid of 200x200 neurons with uniform activation levels after 100 network iterations. A regular local connectivity is used with maximum length 3 (A) and 10 (B). Simulations in (C,D) use the Gabor receptive fields and horizontal oscillator connectivity from the correlation statistics. The input images are artificial stimuli of two collinear aligned bars (C) and of a dashed circle (D) and are shown as an inset in the upper left corner. The shown phase maps in (C,D) are the circular mean of the phase variables after 20 network iterations weighted by the activation of the corresponding features.

hand, in these simulations the connectivity is learned based on natural stimuli and not hand crafted. Hence, it demonstrates that these Gestalt properties are learned from the statistics of natural stimuli.

We next evaluate the concept of binding by synchrony also on natural visual scenes. All following simulations in this paper use color images from the LabelMe database (Russell et al., 2008) and either the Gabor filters or the autoencoder filters to generate the activation levels for the network. An example of a suburban scene is shown in **Figure 3.8A** with the corresponding human labeled segmentation masks in **Figure 3.8B**. We use the time constant $\tau = 1/3$ for the simulations based on Gabor filters and $\tau = 1/30$ for the simulations based on autoencoder filters. These values were chosen such that per iteration of the classical Runge-Kutta solver the phase of not more than 1% of all neurons changes more than $\pi/2$. The units of these time constants are arbitrary because our model of coupled phase oscillators describes the change in phase independent of the oscillation period. Examples of the resulting phase maps are shown in **Figure 3.8C** for Gabor activations and **Figure 3.8D** for autoencoder activations. The phase maps of simulations using autoencoder weights are blurred compared to the Gaborfilters because the peak of the receptive fields are not necessarily centered within the convolutional weight matrix, leading to shifts in visual space between different feature maps at segment boundaries. Yet in both examples an intuitive segmentation of the original can be recognized again in the distribution of phase values. We see a constantly increasing phase synchrony in labeled segments. This example suggests that high-level image objects are likely to synchronize to a coherent phase.

Evaluation of Phase Maps

We evaluate the simulated dynamic phase maps and compare them with human labeled binary segmentation masks of high level image objects from the LabelMe database. We begin with an evaluation of the resulting phase maps independently from the labeled image masks to show global properties of the coupled phase oscillator model and the influence of the number of horizontal connections (section 3.1). This is followed by an evaluation of the phase synchrony within labeled segments with respect to the surrounding of the segments (section 3.1). Finally a local evaluation of the phase maps at the boundaries of labeled segments is presented (section 3.1).

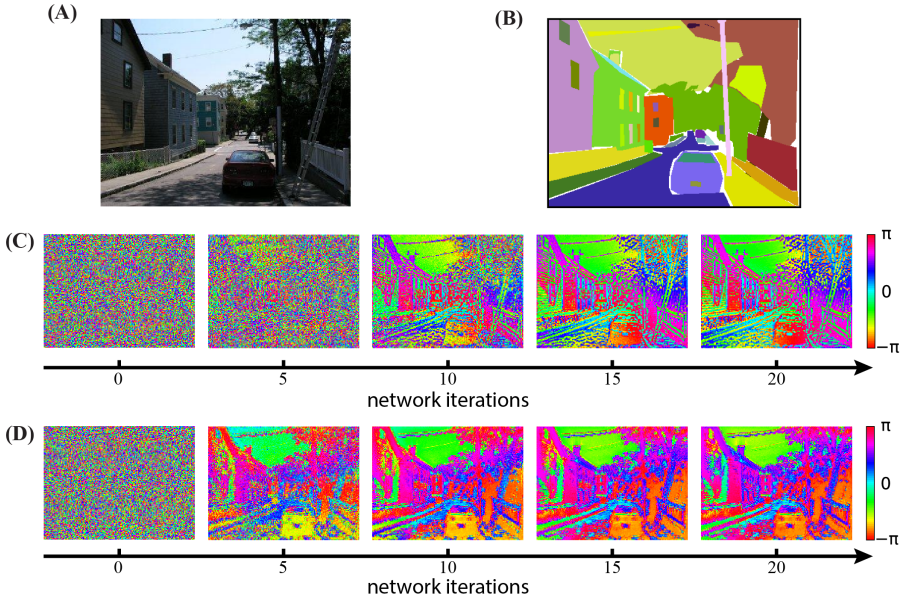


FIGURE 3.8: **Phase simulations of a natural image of a suburban scene.** (A) A natural image from the LabelMe database is used as the input to generate neuronal activation maps. (B) The LabelMe images are accompanied by overlapping segmentation masks of labeled image regions. (C,D) The circular mean of the phase maps evaluated at different network iterations. Gabor filters were used in (C) and autoencoder filters in (D).

Phase Synchrony Segmentation and binding of neurons in the network can only be achieved if the phase variables are not random but also not completely synchronized. Therefore we will first evaluate the local phase synchrony independent of segments in the image. We define the synchrony in a population M of neurons as

$$p_M = \left| \frac{\sum_{m \in M} g_m \cdot e^{i\varphi_m}}{\sum_{m \in M} g_m} \right|, \tag{3.22}$$

where M is defined as a set of 3-dimensional indices describing the position of the neurons.

In this section we analyze the simulation shown in **Figure 3.8** in more detail

and evaluate how the number of synchronizing and desynchronizing connections effects the phase synchrony. We evaluate the local phase synchrony at image position (x, y) for a certain radius r by calculating $p_{M_{x,y,r}}$ for neurons at positions

$$M_{x,y,r} = \left\{(\tilde{x}, \tilde{y}, k) \middle| (x - \tilde{x})^2 + (y - \tilde{y})^2 < r^2, (x, y) \in \mathbb{N}^2, k \in \{1..K\}\right\}, \tag{3.23}$$

where K is the number of feature maps. We average this quantity over all possible image positions (x, y). This mean local phase synchrony is shown in **Figure 3.9** for simulations using different number of connections, different iterations and for different radii r.

When the network has reached a steady state, the mean local phase synchrony depends on the number of synchronizing and desynchronizing connections (**Figure 3.9A,D**). The number of synchronizing connections increases the average local phase synchrony. In contrast, the number of desynchronizing connections can increase or decrease the average local phase synchrony depending on the number of synchronizing connections. At first sight, this may be counterintuitive. In the case of few synchronizing connections, the desynchronizing connections repel the associated phase variables from each other. This ultimately leads to a clustering in the circular phase space evoked by desynchronizing interactions. In the case of more synchronizing connections, the main force driving the network are attractor states and therefore desynchronizing connections decrease the overall phase synchrony.

The phase synchrony in the steady state condition increases with the ratio between synchronizing and desynchronizing connections up to a ratio of 16 times more synchronizing than desynchronizing connections (**Figure 3.9B,E**). Interestingly, the phase synchrony in the steady state condition decreases again in simulations with more than 800 synchronizing connections and very few desynchronizing connections. During the transient phase a very low or high ratio leads to a faster convergence to a more synchronized state. The slowest convergence is achieved at the cases with 4 times more desynchronizing connections or when the number of synchronizing and desynchronizing connections is balanced.

The phase simulations show synchronization behavior at a large variety of different spatial scales (**Figure 3.9C,F**). The level of synchrony at the steady state

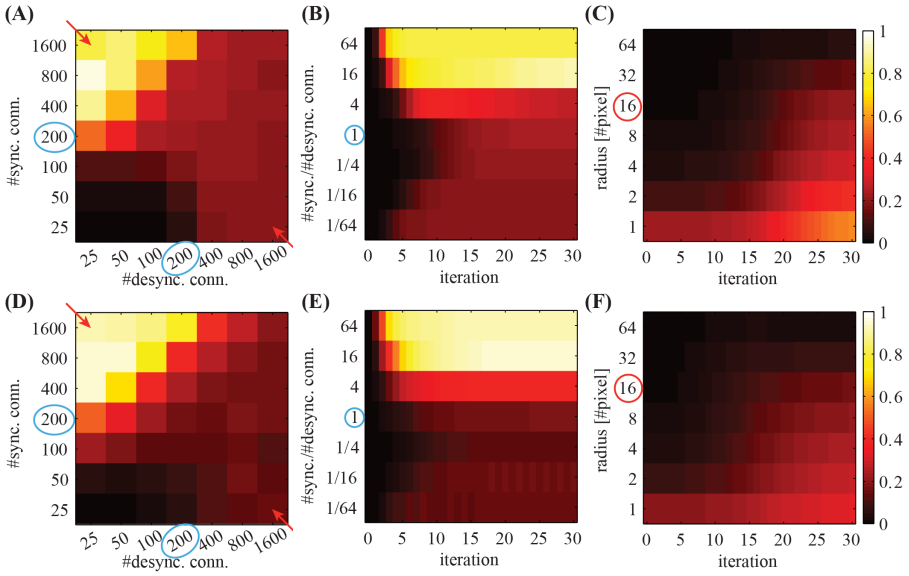


FIGURE 3.9: **The averaged local phase synchrony in circular image regions for different simulation parameters.** All evaluations are based on the activation levels obtained from the image shown in Figure 3.8A. Simulations in the top row (A,B,C) are based on Gabor weights; simulations in the bottom row (D,E,F) are based on autoencoder weights. Blue circles indicate the standard parameters for subsequent evaluations. Colorbars of all panels are the same and shown on the right. The panels in the left column (A,D) show the phase synchrony after 20 iterations for different number of excitatory and inhibitory phase connections per neuron. The panels in the center column (B,E) show the phase synchrony for different ratios of excitatory to inhibitory connections as a function of network iterations. These ratios correspond to the diagonal elements marked with red arrows in panels (A,D). And the shown time course of the average phase synchrony values are from the same simulations. In the right panels (C,F) the phase synchrony is shown for different sizes of the local circular region of the evaluations. The red circle indicates the radius which was used in the evaluations shown in the other panels.

decreases for increasing radius of the phase synchrony evaluation. At all spatial scales the time to reach the steady state synchrony level is roughly the same. Only very localized regions over 1-2 pixel distances show a slightly faster convergence to the final phase synchrony level. When not otherwise stated we select in all simulations and evaluations an intermediate parameter range with balanced synchronizing and desynchronizing connections leading to rich dynamics. These standard parameters are marked with blue circles in **Figure 3.9**.

Segmentation Index The dynamic binding and segmentation of the simulated phase maps of natural images are evaluated using hand labeled segmentation masks. Here a baseline is necessary to accommodate for the higher probability of synchronization between neurons that are close by. Consequently we use the labeled image masks on the corresponding simulated phase maps and compare them to a baseline using the same image masks on simulations of different nonmatching images.

The segmentation masks in the LabelMe database are specified as polygons on the images that are initially reduced in our simulation to a resolution of 400x300 pixels. The convolutional forward projections lead to a further reduction in the feature representation to a grid of 200x150 pixels. Therefore we restrict the evaluations of the phase maps to segmentation masks which contain at least as many pixels as the specified patch size of the forward projections (6x6 neurons corresponding to 12x12 pixels in the input image). In addition, segments occupying more than half of the respective images are excluded to allow evaluations against a baseline synchrony of the surrounding regions. The range of labeled segments which is used in our evaluations is shown as a horizontal bar in **Figure 3.10**. Only in evaluations where the segment sizes are explicitly stated, we also evaluate these otherwise excluded very small and very large segments.

The number of labeled segments in the database decreases for larger segment sizes (**Figure 3.10A**). Yet the total area occupied by segments in the different bins increases for larger segment sizes (**Figure 3.10B**). Therefore when applying labeled masks to nonmatching images small segments are highly likely to fall into large segments where a large number of tangential connections is functionally active. Consequently the phase synchrony within labeled segments is

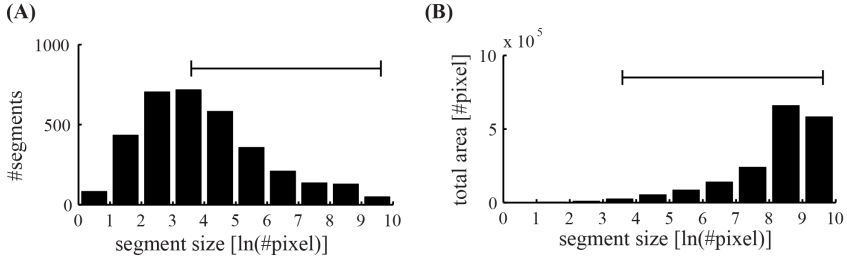


FIGURE 3.10: **Statistics of labeled image segments.** (A) The histogram of evaluated segments from the LabelMe database for different segment sizes is shown. (B) The total area occupied by the segments in the corresponding bins. The range of segment sizes (36-15000 pixels) that are used for subsequent evaluations are marked with a horizontal bar in (C,D).

not a sufficient baseline for an unbiased comparison with simulations of non-matching images. Therefore we need a baseline to control for the unequal distribution of segment sizes and their occupied region in the images.

To accommodate for the statistics of segment sizes in the evaluation of the matching and nonmatching natural scenes, we define a segmentation index (**Figure 3.11**) that sets the phase synchrony in segments into the context of the surrounding neurons. Concretely, the segmentation index evaluates how the phase of neurons inside of segments is more or less synchronized compared to the synchrony of random neurons inside and outside of the segment. The neighborhood N of a segment Q is generated using a diamond shaped grow operation on the segmentation mask repeatedly until the number of neurons in N is doubled compared to the original segment Q. Therefore N is the union of the segment Q and the surrounding R of the segment (Q and R are annotated in the example shown in **Figure 3.11**).

We calculate the phase synchrony values p_{Q_j} and p_{N_l} for random subsets $Q_j \subset Q$ and $N_l \subset N$ where $j, l \in \{1, ..., 100\}$ and $Q_j, N_l \in \mathbb{N}^{1000}$. We define the segmentation index of segment Q as the difference between the mean synchrony within the segment Q to the mean synchrony in the neighborhood $N = R \cup Q$:

$$\kappa(Q, N) = \langle p_{Q_j} \rangle_j - \langle p_{N_l} \rangle_l. \tag{3.24}$$

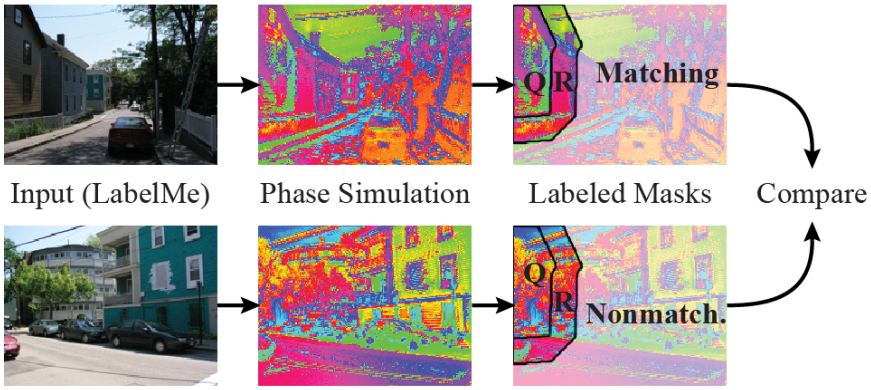


FIGURE 3.11: **Evaluation using hand labeled image masks.** The evaluations compare the segmentation index of matching simulations and segmentation masks (top row) to a baseline of nonmatching simulations and segmentation masks (bottom row). The images from the LabelMe database (left column) are processed using the forward projections. The resulting features are used to simulate the phase of the coupled neural oscillators (middle column). The segmentation index of these phase maps are then evaluated using the segmentations masks from the LabelMe database (right column). The evaluation of the house in the top left is here shown as an example. The segmentation index compares the phase synchrony in the hand labeled region of the house (Q) to a baseline phase synchrony within the neighborhood ($Q \cup R$).

The segmentation index increases over simulation iterations for matching and nonmatching masks and images (**Figure 3.12A**). The matching conditions have a steeper ascent and reach a higher segmentation index compared to the nonmatching conditions. The difference between the matching segmentation index and the nonmatching segmentation index increases for both simulations using Gabor weights and autoencoder weights (**Figure 3.12B**). The simulations using regular Gabor receptive fields show larger differences between matching and nonmatching segmentation indices compared to the autoencoder weights. The ratio between matching and nonmatching segmentation indices is roughly the same for both types of receptive fields. This demonstrates systematic binding in the phase maps of matching segments.

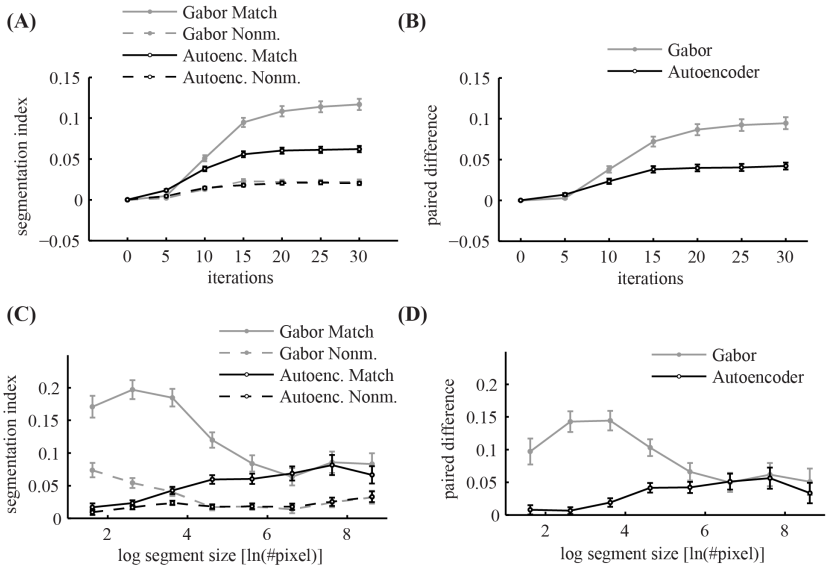


FIGURE 3.12: **Segmentation index.** The mean segmentation index is shown as a function of network iterations averaged over all segments with more than 36 pixels in the top panels (A,B). The segmentation index is shown as a function of different segment sizes after 20 network iterations in the bottom panels (C,D). The panels on the left side (A,C) show the evaluations for matching images (solid lines) and nonmatching images (dashed lines) individually. Panels on the right side (B,D) show the paired difference between matching and nonmatching evaluations. In all panels the activation levels are obtained using Gabor filters (gray lines) and autoencoder filters (black lines). The errorbars in all panels are 95 % confidence intervals.

An evaluation for different segment sizes individually reveals more differences between the Gabor and autoencoder features. The evaluations of the matching conditions show that the segmentation index increases for larger segments in the case of the autoencoder features but decreases for larger segments in the case of the Gabor features (**Figure 3.12C**). An explanation is that the autoencoder contains more features with low spatial frequencies while the Gabor features are restricted to one specific spatial frequency.

The paired difference between matching and nonmatching evaluations shows that the Gabor filter and the autoencoder have roughly the same performance for large segment sizes (**Figure 3.12D**). For small segment sizes the autoencoder has a decreased segmentation performance. One possible explanation might be that the receptive field weights are not centered (compare **Figure 3.3**) and therefore different feature neurons might be slightly misaligned relative to the hand labeled segmentation masks, which are defined as polygons with arbitrary precision on the image.

Overall the results show a significant difference between the matching and the nonmatching segmentation indices for all evaluated segment sizes. The paired difference between the matching and the nonmatching conditions increases as the simulation of the randomly initialized phase variables slowly converges to a state with clusters in the circular phase space. After about 20 network iterations the paired difference in the segmentation index reaches a high plateau. Therefore the coupled phase oscillator model achieves a stable segmentation of the natural image scenes with a coding of binding by synchrony.

Segment Boundaries To evaluate how well the phase maps segment different labeled regions at their borders we calculate a metric at random locations of segment boundaries. We sample 50 random locations from all boundary lines of the segments in each simulated image from the LabelMe database. At these locations we use the angle of the segment boundary to divide a local region into two semicircles with a radius of 10 pixels such that one half lies approximately within the segment and the other half outside of the segment (**Figure 3.13A**). The mean phase difference between both semicircles decreases over simulation time (**Figure 3.13B**). The paired difference between the phase difference in matching compared to nonmatching images shows that the phase difference over matching segment boundaries is significantly larger (**Figure 3.13C**).

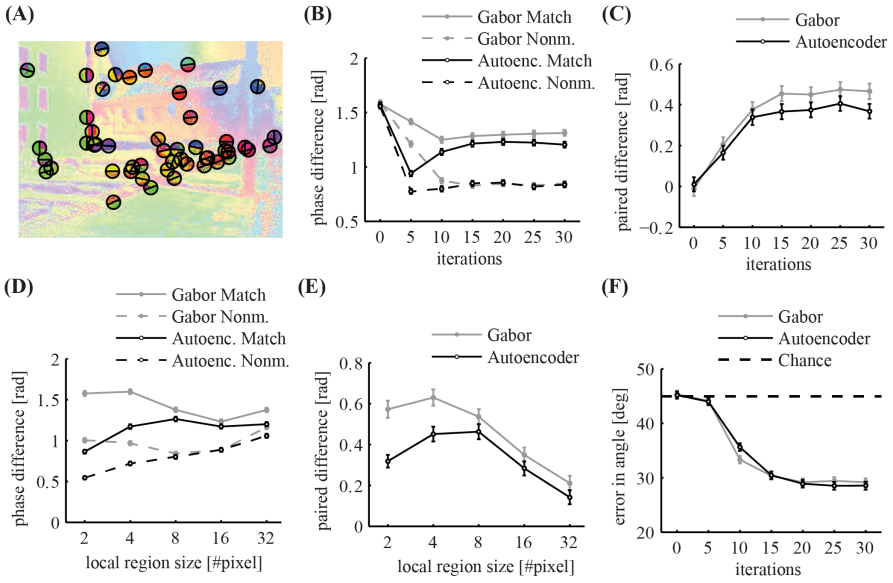


FIGURE 3.13: **Local evaluation of the phase segmentation.** Results were obtained using Gabor filters (gray) and autoencoder filters (black). (A) Illustration of the randomly selected locations on segment borders and the corresponding semicircles as described in the main text. (B) The local phase difference at random segment border locations of matching images (solid lines) and nonmatching images (dashed lines). (C) The paired difference between the local phase differences evaluated on matching and nonmatching images. (D) The mean local phase difference as a function of different sizes of the local circular regions over which the phase is evaluated. (E) The paired difference between matching and nonmatching images. (F) The mean error in the estimated angle of segment boundaries. All errorbars are 95 % confidence intervals.

The evaluation of the phase difference as a function of the size of this circular region shows that the segmentation performance using autoencoder features decreases for very small regions (**Figure 3.13D-E**). This might be due to the above described misalignments between the learned receptive field centers. For very large evaluation regions the performance decreases for both receptive field types because the circular regions are likely to extend beyond the hand labeled segment regions.

It is possible to evaluate the segmentation performance of the dynamic binding maps without the need for a baseline on nonmatching images if we use an unbiased performance estimator with a clearly defined chance level. Therefore we measure how well the phase map can predict the angle of the borders of segmentation masks. We use the phase variables at randomly sampled locations on segment boundaries (**Figure 3.13A**) and compute the image direction with the largest change in the phase variables. We define the local variance in phase at image position (x, y) as

$$\vartheta_{x,y} = 1 - \frac{1}{5 \cdot K} \cdot \left| \sum_{k=1}^{K} e^{i\varphi_{x,y,k}} + e^{i\varphi_{x-1,y,k}} + e^{i\varphi_{x,y-1,k}} + e^{i\varphi_{x+1,y,k}} + e^{i\varphi_{x,y+1,k}} \right| \tag{3.25}$$

where the sum is over all $k \in \{1..K\}$ feature maps. We use the structure tensor of the local variance in phase to estimate the principal directions. To compute the structure tensor we use a Gaussian window function with a standard deviation of 3 pixels and the second order central finite difference of the local variance in phase. The eigenvector of the structure tensor gives an estimate of the border direction of the segmentation mask. The evaluation of the phase maps shows that the mean error in the estimation of the boundary angles decreases over simulation time (**Figure 3.13F**). A minimum is reached after around 20 network iterations with an error of approximately 28° in comparison to the chance level of 45°. This demonstrates that the phase gradient systematically aligns itself orthogonal to the segment boundaries.

Discussion

Here we investigate the concept of binding by synchrony, as has been previously studied with abstract stimuli, in the context of unsupervised learning and

natural stimuli. The model consists of coupled phase oscillators with a connectivity based on natural image statistics. Specifically, the correlation of neuronal activity governs the structure of local horizontal connections in the network. Hence the connections are not constructed according to a heuristic or intuition, but solely data driven. Therefore we can expect it to generalize well to other cortical areas. We show that the sampled sparse connectivity based on positive correlations in induced activations by natural stimuli exhibits small-world properties. We hypothesize that the small world property is a signature of Gestalt laws in the form of regular local correlations (objects) that can be flexibly combined on a global scale. We show that these horizontal connections influence the dynamics of the phase variables such that an effective coding of contextual relationships between active neurons is implemented by phase synchronization. Therefore our results reveal that the concept of binding by synchrony is viable for natural stimuli.

The evaluation of phase synchronization as a code for grouping and segmentation utilizes hand labeled image segments, corresponding to high level objects, as ground truth. The evaluations reveal that the phase maps are binding active neurons together if they encode different attributes of the same stimulus. It follows that the phase variables are coding global stimulus attributes in contrast to the coding of local stimulus attributes by the rate variables. The coding of these global contextual relationships is not directly influenced by the rate variables but only by their indirect modulation of the phase interactions. Furthermore we illustrate that discontinuities are formed in the phase maps at the borders of segments and that these discontinuities can predict the orientation of segment boundaries. Therefore our results suggest that the segmentation driven by bottom up dynamical processes using natural image statistics matches to a certain degree the top-down labeling of abstract image objects.

Our study connects three different subject areas: natural image statistics, dynamical models of neural networks and normative models of sensory processing. In the following we will discuss the motivations and implications of our study from each of these perspectives.

Choice of Natural Stimuli

The choice of "natural" stimulus material is not as obvious as it might seem. A more natural choice from a biological perspective would be to use stimulus material generated by a moving agent. For example videos from a camera mounted to a cat's head were used previously to analyze the spatio-temporal structure of natural stimuli (Kayser, Einhäuser, and König, 2003). A similar setup from a human perspective is also possible (Açik et al., 2009). But time variant stimuli require more computational resources and the high number of horizontal connections in our simulations is computationally expensive although it is implemented as a vectorized operation. In addition, the analysis of the phase segmentation maps would be more difficult in the case of moving stimuli because of the unknown time lag between stimulus onsets and the resulting dynamic phase maps. Therefore we decided to not use videos as stimulus material in the present study.

Differences in eye movements given different stimulus classes might also play a role in shaping the statistics in the visual input received by the primary visual cortex. There might be important interactions between saccadic eye movements and the dynamics of the horizontal connections in the visual cortex. One could simulate saccadic movements on static images using saliency maps and use the resulting images for the feedforward processing in our model. But as with moving stimuli in general it would complicate the analysis and would not contribute directly to the understanding of the central questions of binding by synchrony.

The LabelMe database provides a large set of only static images. It has the advantage that the images are accompanied by labeled region masks of well defined objects. These high level labeled masks are often overlapping in the case of part-based segmentations of objects. The segmentation evaluation is tricky in the case of occluded objects. But the LabelMe database allows us to investigate the relationship between natural image statistics and the coding of high level image concepts. Therefore we think it is a reasonable choice to use this database in our study.

Biological Plausibility

As with most computational neural network models we have to ask ourselves in how far it is biologically plausible. To advance our knowledge about the underlying computation principles in the cortex, it is always a good choice to model only the level of detail which is necessary to explain the phenomena under investigation. Thereby we assure that the abstraction level of the model is as good as possible although it is very likely that some mechanisms below the level of detail modeled here play an important role in synchronization phenomena. We implement in our simulations the influence of correlated neuronal activity on large time scales to the network connectivity. Based on these connections we show how the dynamics on fast time scales can code for segmentation and binding. Therefore we have to model the behavioral learning time scales (>days) to capture the natural image statistics and the dynamical network time scales (<seconds) simultaneously. Therefore we consider the chosen network architecture of segregated rate and phase based coding suitable to investigate the role of correlated neuronal activity on the network dynamics and relational coding by synchronization.

The Kuramoto model restricts the dynamical interactions between coupled oscillators to a scalar phase variable. Breakspear, Heitmann, and Daffertshofer, 2010 review this simplified model of coupled phase oscillators in the context of models of complex neurobiological systems. They find that it captures the core mechanisms of neuronal synchronization and a broad repertoire of rich, non-trivial cortical dynamics. Studies of the Kuramoto model mostly focus on regularly defined phase interactions without a separate network variable representing the activation levels of the oscillator neurons. This allows using mean-field approximations to further simplify the analysis of the Kuramoto model. In contrast, our study focuses on the simulation of heterogeneous connections which are modulated by heterogeneous activation levels induced by natural stimuli. Therefore our simulation model is more similar to the diverse activations and connections found in biological neural systems but this comes with the drawback that a mean-field approximation is not warranted.

In principle two biological interpretations of the coupled phase oscillator model are possible. A conservative standpoint is an interpretation as a neural field model in which each network unit of our simulation represents a functional

module, i.e. a cortical column, which is comprised of many biological neurons. In this case the phase variables would represent the average phase of a set of biological neurons, i.e. the phase of the local field potential. A second possible more fine-grained interpretation in which the phase oscillators represent individual biological neurons might seem far-fetched and oversimplified on first sight. Nonetheless the interpretation of the phase variables as spike timings might give further ideas about possible extensions of our proposed model. In this interpretation the oscillators represent the limit cycles of the dynamics of spike generation of biological neurons. The sinusoidal interaction function can then be related to an integral over the phase response function of a spiking neuron (Sturm and König, 2001). Furthermore the spike interpretation could motivate the introduction of conduction delays in our model. This in turn might further allow studying spike-timing dependent plasticity in the context of a normative model.

Certainly, there are many phenomena that can only be modeled by more detailed spiking neuron models. For example spike-timing dependent plasticity could only be modeled with the phase oscillator model if we assume regular oscillatory firing but not in the case of irregular firing. For example, the ability of self-organizing recurrent networks (SORN) to learn spatio-temporal structures in the input depends on spike-timing dependent plasticity and irregular firing (Lazar, Pipa, and Triesch, 2009). Similarly, Buonomano and Maass, 2009 showed that spatiotemporal processing of natural stimuli can emerge from the dynamics of 'hidden' neuronal states, such as short-term synaptic plasticity. Irregular firing is also needed for synfire chains of successively activated neural assemblies to explain the physiological measurements of spike patterns recurring with millisecond precision (Abeles, 1982). However it might be possible to simulate some properties of synfire chains if we add more hierarchical layers and phase conduction in the feed forward projections in our model. Kumar, Rotter, and Aertsen, 2010 analyzed the coexistence of firing rate propagation and synchrony propagation in feed forward networks. Last but not least, self-organized criticality and cortical avalanches (Beggs and Plenz, 2003) can probably only be modeled with more detailed spike-based neuron models because the phenomenon requires a dynamical system of more complex coupled oscillators.

There are also other dynamical models of neural networks that were analyzed in the context of scene segmentation (Tononi, Sporns, and Edelman, 1992).

Wang and Terman, 1997 described the local excitatory global inhibitory oscillator network (LEGION), which is comprised of units described by two differential equations that explicitly model a stable periodic orbit alternating between two phases with rapid transitions between them. This model has the advantage that fast synchronization of the coupled oscillators is possible. But it simulates each neuronal oscillation on a fast timescale and the synchronization of a population of neurons is only visible at certain simulated time points. In contrast, our phase model simplifies the phase plane to a continous phase variable averaged over many oscillatory periods, so that the phase relationships between all pairs of neurons is explicitly represented at all simulation time points. Another difference is that the implementation of LEGION involves many discontinuous operations to reduce the computation time. These discontinous operations prevent a normative model approach with optimizations using gradient descent. The full continuous dynamics in our model allows further optimizations of the horizontal connectivity using gradient descent methods.

In our model the forward connections are computed once and are then fixed during the phase simulation of horizontal connections. This is a very simplified model compared to the ongoing simultaneous processing of afferent and recurrent inputs in the cortex. But it is compatible with the fact that self sustained activity in the cortex can be measured also in the absence of stimulus inputs. Furthermore, computational models of cellular and network behavior support the conclusion that the cortical network operates in a recurrent rather than a purely feed-forward mode (Mariño et al., 2005). Therefore, it makes sense to simulate the lateral interactions decoupled from the time scale of forward projections that generate the activation levels.

We use the correlated neuronal activation levels as the probability to form horizontal intralayer connections. It was shown that the measured horizontal connectivity in the visual cortex of cats is indeed proportional to the correlation between receptive field wavelets in image statistics (Betsch et al., 2004). Our choice to use a sparse connectivity pattern instead of full connectivity with heterogenous connection strengths was initially intended as a computational shortcut to allow large-scale simulations. This sparse connectivity is in line with biological horizontal connectivity and reveals interesting properties that deserve further investigation. In the brain the binding of stimulus representations has to be distributed over many cortical areas. It was shown with graph theoretic measures that the sparse connectivity within the cortex is organized

in hubs and shows properties of small-world networks (Sporns et al., 2004). One can speculate that this allows binding by temporal structure even between stimulus representations over distant cortical regions. Also in our network model the sampled sparse connection patterns generated from correlated neuronal activity were shown to have small-world properties in the case of synchronizing connections. Accordingly, we see in our network simulations fast synchronizations of distant neurons that are not directly connected. And in future studies our model could be extended to simulate even synchronizations between different cortical regions.

In the cortex a wide range of oscillatory frequencies at different spatial scales occur with cross-frequency couplings. This is highly prominent in different sleep stages (Belluscio et al., 2012) and plays an important role in memory encoding (Friese et al., 2012). Our model is highly simplified in the sense that all neurons are assumed to have the same oscillatory natural frequency. We simulate only horizontal connections between neurons with similar physiological properties which are operating in the same dynamical regime. In this context, the assumption that all active neurons are close to a similar dynamical limit cycle seems reasonable. In future work, several cortical rhythms could be implemented using several phase variables per neuron. One can conceive different algebraic structures which could efficiently represent cross-frequency couplings in the cortex. This would allow investigating fractal binding at different abstraction levels and segmentation at different scales.

In summary, the architecture of our model captures many important aspects of biological neural networks. In particular, it models the dynamical properties used for contextual coding and the unsupervised learning of statistics in natural stimuli. At the same time, our model keeps the simplicity required for the analysis of the network dynamics and allows relatively simple evaluations of the resulting phase relationships.

Comparison with other Normative Models

In recent years the abstraction from complex differential equations describing biological neural networks to normative models of rate-based sensory processing improved our knowledge on the underlying computational principles of the cortex (Olshausen and Field, 2005). Unsupervised learning of the inherent

statistics in the sensory input seems to be one of the main mechanisms governing the structural connectivity between neurons in low level sensory areas of the cortex (Olshausen, 1996; Wiskott and Sejnowski, 2002; Körding et al., 2004). On the other hand relatively few studies have investigated the relationship between unsupervised learning using correlated neuronal activity and the coding of contextual relationships through binding by synchrony. In this section we describe differences and similarities between our model and other normative models of sensory processing in the brain.

Wyss, König, and Verschure, 2006 and Franzius, Sprekeler, and Wiskott, 2007 show that rate-coding neurons form a hierarchy of processing stages resembling the ventral visual pathway. These studies use optimization functions of optimal stability and decorrelation while exposing the network to natural stimuli. Although these models provide important insights into the information processing mechanisms in the cortex, they don't take into account the processing of contextual information and lack an implementation of relational coding between different features. In a similar way to these studies, we use the statistics of natural stimuli not only to learn feature representations but also to explain relational coding in the context of binding by synchrony. This approach could allow combining multi-scale image segmentation and object recognition into a hierarchical neuronal network model. A prerequisite for analyzing the segmentation by synchrony in a hierarchical network is an unsupervised learning of the feed-forward connections to generate the activation levels for higher network layers. We have shown that the proposed segmentation by synchrony works with receptive fields obtained from convolutional autoencoders, which can be stacked to obtain the forward and backward connections within a hierarchy. This allows a completely unsupervised learning of feed-forward, feed-back and intralayer connections using natural image statistics. Binding and extraction of features can be accomplished simultaneously within the hierarchy.

Biologically inspired autoencoder models were shown to be efficient for unsupervised learning of receptive fields by minimizing the reconstruction error of the input (Coates, Lee, and Ng, 2010). Complex valued autoencoders have similar to our model 2 variables per network node (Baldi and Lu, 2012). To our knowledge the available publications investigating complex valued autoencoders focus mainly on the aspect of learning compressed representations of complex valued inputs. They do not directly address the biological motivation

of binding by synchrony. They are usually strictly defined on the typical complex algebra and are not described by a differential equation which corresponds to coupled oscillators. The formalism of complex valued autoencoders might be adapted to allow further abstractions of our model. This could support our understanding of the underlying computational principles of visual grouping and segmentation.

A very different and novel approach of coding contextual informations in autoencoder networks are mean-covariance restricted Boltzmann machines (Ranzato and Hinton, 2010). In these models latent hidden factors are used to efficiently represent the contextual information in the input in addition to the usual representation of pixel means in standard models of restricted Boltzmann machines. It was shown that the model can efficiently code pixel covariances in analogy to complex cells and pixel means in analogy to simple cells. However the coding of contextual information in these models is limited to pair-wise interactions in the input layer. Therefore this kind of generative model can capture only a linear combination of second order statistics so that contextual interactions between a large group of neurons is only possible through direct connections. In contrast, the grouping in our model is a dynamic process in which interactions between neurons are possible without a direct connection between them but through intermediate neurons. The reason is that our model uses a dynamical system approach with recurrent connections in contrast to probabilistic modeling of forward and backward connections.

Some mathematical theories of cortical processing mechanisms also take the contextual information into account. For example the free energy principle (Friston, 2010) and the theory of coherent infomax (Kay and Phillips, 2011) explicitly incorporate the context into single-variable local processors in the network. In contrast, the model presented in this paper takes the context into account in a separate phase variable, which codes relational properties similar to the dynamics on fast time scales in biological neural networks. Thereby our simulation allows to model higher order relational structures with a limited number of horizontal connections. In contrast, in the mathematical formalization of coherent infomax the contextual field input is assumed to be integrated into a single variable output of a local processor in the network. Thereby it doesn't allow implementing higher order relations between many local processors if the computational resources are limited. This limitation is of course

only a matter of the used mathematical formalism and doesn't affect the general explanatory power of the free energy principle or the theory of coherent infomax. Therefore, in a broader sense our simulation model could be seen as an approximate implementation of these abstract concepts, although we use a biologically motivated architecture instead of a probabilistic derivation.

Our study combines aspects of these normative models of sensory processing and of detailed models of dynamical neural networks. We use only the statistics induced by natural images to learn unsupervised the forward and tangential phase connections. The supervised labeled segmentation masks are only used to evaluate how phase synchrony corresponds to a relational coding in the neural representation. Hence, the concept can be phrased completely in the form of a normative model. In future work, we plan to further formalize the model and conceive more complex learning rules for the phase interactions. These learning rules could replace the sampling of sparse connections from the correlation of activation by a more biologically motivated rule. For example, one could develop learning rules based on spike-timing dependent plasticity if phase delays are incorporated in the interactions of the network. This would additionally allow modeling phase locking between neurons and coding of syntactic relations in the network. These extensions to our model could provide new insights into the computational principles underlying higher order cognitive processes.

Conclusions

Our study revealed that the concept of binding by synchrony is viable in the context of unsupervised learning using natural stimuli. We show that the structural connectivity based on correlated activity leads to relational coding in a neural network model of coupled phase oscillators. The presented novel evaluation methodology for image segmentation revealed that the phase of neurons code global stimulus attributes. This strengthens the evidence that phase synchronization plays a key role to coordinate the spatially distributed information processing in the cortex. One could further speculate on how higher level coordination and binding between cortical areas might evolve from unsupervised learning based on correlated neuronal activity.

Disclosure/Conflict-of-Interest Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgement

The authors would like to thank Robert Märtin for his valuable comments and helpful suggestions.

Funding: This work was funded by the DFG through SFB 936 Multi-Site Communication in the Brain.

3.2 Chapter Summary

In this chapter, I presented a model of normative learning from natural images that we extended with phase variables to model binding by synchrony. We showed that coding contextual information with oscillatory interactions is feasible in the setting of unsupervised learning of distributed neural representations. Importantly, we not only learned rate-based sensory representations from natural sensory statistics, but also the horizontal intracortical connectivity was learned unsupervised from the activation statistics of the corresponding nodes in the network.

Here, the forward connections, that give rise to the distributed representations of the sensory inputs, were learned by an autoencoder with three optimization terms. The minimization of the reconstruction error was supplemented with an optimization term for the sparseness of the hidden layer activations and a weight decay term to regularize the learning process. The resulting neural representations of the sensory inputs were shown to be sparse, meaning that for a given input image most neurons are inactive while only few are actively encoding the visual scene. This sparseness of the activation levels was very important for the following contextual coding by phase variables. One can speculate that the activation sparseness is important because it induces a new topological network structure for the horizontally coupled oscillator network depending on the current sensory inputs. Therefore, the simulation of the coupled oscillator interactions can be seen as taking place within a specific unique instantiation of a new network topology that is induced by the current sensory input.

The unsupervised learning of the horizontal connections through which the oscillatory interactions take place is an integral part of this model. For these connections we assumed that anti-correlated neural activity leads to negative connection strengths, which would result in the desynchronization of the coupled phase oscillators. Specifically, negative connection strengths would push the two interacting nodes into the synchronized state, while positive connection strengths would push the nodes into an anti-phase constellation. Interestingly, we found that the negative connection weights not only decrease the overall synchrony in the network if all nodes are synchronized, but importantly, that they also increase the overall synchrony if all nodes have random

phases. Therefore, these negative connections are important for achieving and maintaining a homeostasis of the overall synchronization state in the network, resulting in the binding of some but not all active neurons.

As a result we can interpret this manuscript in the light of Figure 1.1 that was presented in the introduction. The activation variables in the network implement the functional activation (center circle in figure) on the medium time scales that are produced by the perception of the environment and could be further used to execute actions in the environment (top circle in figure). The autoencoder learns the structural connectivity (left circle in figure) on slower time scales based on repetitive functional activations. On the other side, the functional activation also induces very fast dynamics in the coupled oscillator network (right circle in figure). Importantly, these phase oscillators interact in a topological space that is constrained by the functional activation, and therefore is dynamically organized without the necessity to change the underlying anatomical structural connectivity.

In summary, this chapter showed how learning from activation statistics and oscillatory interactions allow the dynamic organization in the microscale circuitry, while the previous chapter demonstrated how large scale cortical interactions are mediated by the anatomical structure. In the next chapter, these interactions are further investigated from an even broader perspective of behavioral and cognitive functions. In this regard, I will present two manuscripts: The first manuscript will analyze the behavior of subjects in the context of multisensory integration leading to further insights into the dynamics of cortical multi-site interactions. Subsequently, the second manuscript will investigate cortical oscillatory interactions in the context of associative memory consolidation and retrieval.

Chapter 4

Relation to Behavior and Cognitive Functions

4.1 Bayesian Alternation During Tactile Augmentation

This section was published as a peer-reviewed article: Caspar M. Goeke, Serena Planera, Holger Finger, and Peter König (2016). "Bayesian Alternation during Tactile Augmentation". In: *Frontiers in Behavioral Neuroscience* 10, p. 187. ISSN: 1662-5153. DOI: 10.3389/fnbeh.2016.00187. URL: http://journal.frontiersin.org/article/10.3389/fnbeh.2016.00187

Abstract

A large number of studies suggest that the integration of multisensory signals by humans is well-described by Bayesian principles. However, there are very few reports about cue combination between a native and an augmented sense. In particular, we asked the question whether adult participants are able to integrate an augmented sensory cue with existing native sensory information. Hence for the purpose of this study, we build a tactile augmentation device. Consequently, we compared different hypotheses of how untrained adult participants combine information from a native and an augmented sense. In a two-interval forced choice (2 IFC) task, while subjects were blindfolded and

seated on a rotating platform, our sensory augmentation device translated information on whole body yaw rotation to tactile stimulation. Three conditions were realized: tactile stimulation only (augmented condition), rotation only (native condition), and both augmented and native information (bimodal condition). Participants had to choose one out of two consecutive rotations with higher angular rotation. For the analysis, we fitted the participants' responses with a probit model and calculated the just notable difference (JND). Then, we compared several models for predicting bimodal from unimodal responses. An objective Bayesian alternation model yielded a better prediction ($\chi^2_{red} = 1.67$) than the Bayesian integration model ($\chi^2_{red} = 4.34$). Slightly higher accuracy showed a non-Bayesian winner takes all (WTA) model ($\chi^2_{red} = 1.64$), which either used only native or only augmented values per subject for prediction. However, the performance of the Bayesian alternation model could be substantially improved ($\chi^2_{red} = 1.09$) utilizing subjective weights obtained by a questionnaire. As a result, the subjective Bayesian alternation model predicted bimodal performance most accurately among all tested models. These results suggest that information from augmented and existing sensory modalities in untrained humans is combined via a subjective Bayesian alternation process. Therefore, we conclude that behavior in our bimodal condition is explained better by top down-subjective weighting than by bottom-up weighting based upon objective cue reliability.

Introduction

Humans sample information from their environment by many senses. In most circumstances (i.e., outside of the lab), behavior is not guided by a single modality but by a combination of several modalities. In the last decade many studies have shown that this process follows optimal Bayesian principles (Ernst and Bülthoff, 2004; Körding and Wolpert, 2004; Körding and Wolpert, 2006). A core concept of Bayesian integration is that perceptional noise (variance) is reduced in multimodal conditions, improving the precision of later decision processes. Many studies concentrated on the combination of visual and haptic cues. Ernst and Banks (2002) showed that visual and haptic information about object sizes are statistically optimally integrated. Extending this idea, Helbig and Ernst (2007) demonstrated optimal integration between vision and touch also for the shape of objects. Similarly, Reuschel and colleagues

showed that visual and proprioceptive information are integrated in a statistically optimal manner for the perception of geometric trajectory (Reuschel et al., 2010). Moreover, several other combinations of senses have been investigated. Battaglia and colleagues found that visual and auditory information are optimally integrated in a spatial localization task (Battaglia, Jacobs, and Aslin, 2003). Frissen and colleagues reported optimal integration between proprioceptive and vestibular information for spatial updating (Frissen et al., 2011). Accordingly, Butler and colleagues argued that visual and vestibular signals are integrated in a Bayesian way for heading estimation (Butler et al., 2010). All in all, there is rich evidence that sensory information from different modalities is integrated following optimal Bayesian statistical principles.

While the concept of Bayesian optimal integration has been confirmed throughout several experimental paradigms, recent studies showing that integration happens only for redundant sensory information, i.e., both signals have to "describe" the same physical property. In this respect, Körding and colleagues demonstrated that the perceived causal relationship of two sensory signals is a prerequisite for sensory integration (Körding et al., 2007). Wozny et al. (2010) provided further evidence, sowing that the majority of their subjects used a probability matching strategy in a perceptual decision task. Furthermore, the integration of two sensory modalities requires a mapping between the two kinds of information. Mapping in this context refers to the cross-modal associations or correspondences of the sensory cues. For instance, there is a certain mapping of how it feels to hold an object in your hand and how it looks like. This association changes with the softness or weight of the object. Importantly, people can learn such a mapping, even if no prior coupling existed before. In particular, Ernst (2007) showed that subjects were able to optimally integrate visual cues (brightness) and haptic cues (stiffness). Similarly, Kaliuzhna and colleagues demonstrated that subjects integrated arbitrary co-occurring self-motion (vestibular) and visual cues (Kaliuzhna et al., 2015). Furthermore, Kuang and Zhang introduced a new visual-olfactory mapping. In their study the researchers linked two different smells to opposite movement directions in a dot movement discrimination task. After establishing such a pairing the presentation of the olfactory cues biased the perception of visual motion direction (Kuang and Zhang, 2014). For a detailed review regarding cross-model mappings also see Ernst (2006).

If new sensory-mappings are optimally integrated without or after very short

training sessions one could ask the question, do humans innately integrate two co-occurring signals? If not, what would be possible alternatives? In 2008, Nardini and colleagues tested the concept of Bayesian optimal integration in a navigation task with three different age groups: children of 4–5 years of age, children of 7–8 years of age, and adults. Interestingly, they found that both groups of children did not integrate optimally between visual and proprioceptive cues but rather alternated between them. In contrast, adults performed the same task in a "Bayesian optimal" manner (Nardini et al., 2008). Similarly, Gori and colleagues reported that integration of vision and touch before 8 years of age is far from optimal (Gori et al., 2008). According to the authors, this was the case even when the dominating sense was made far less precise than the neglected sense. These results provide evidence that the capability of integrating information in a Bayesian optimal way requires several years of experience and is not an inherent property of our brain. More recently, Chen and McNamara tested how people integrate visual and self-motion cues during spatial navigation and found evidence for Bayesian Alternation even for some adult subjects (Chen and McNamara, 2014). Besides Bayesian Alternation, there is of course the possibility that people only use one cue and completely neglect the other. However, in such a case there is no cue combination, or multisensory processing at all. In fact also other recent studies provide evidence for cue alternation behavior (De Winkel et al., 2013; De Winkel, Katliar, and Bülthoff, 2015; Adams, 2016). The general idea behind Bayesian cue alternation is that both cues are used for the task; however, they are never used at the same time. Instead, the subject switches between one and the other cue based on a Bayesian probability selection mechanism. Hence for each trial one or the other cue is selected while the probability for selecting one cue over the other is given by the respective relative weight for each cue. In summary, several studies in the last decade found evidence for cue alternation behavior. To our understanding this deviates from the majority of findings regarding Bayesian optimal integration and needs to be investigated in more detail.

The mechanisms that underlie the transition from cue alternation to cue integration are usually observable only in children or when sensory signals are explicitly manipulated (i.e., adding sensory noise). However, it is unclear what happens when adult subjects are equipped with a new sense (or an augmented sensory-like cue). Are we able to integrate such new information with the cues we receive from our native modalities or do we have to choose and rather

alternate between the two (similar as children do)? In other words, it is most interesting to examine adults' performance when they are provided with a new, augmented sense which they have to combine with information from their existing (native) senses. We specifically ask the question: Is such a process similar or different to the ones observed in children? Throughout this paper we use the term "native modality" to refer to the information mediated by angular rotation through native sensors like the vestibular system, and augmented modality to information mediated by a sensory augmentation device, even as the subjects did not receive a formal training. Angular rotation in our setup was implemented by a rotaing platform on which the subjects were seated, while the augmented information was mediated via tactile stimulations (for details see the Section Experimental Paradigm). Although different combinations of sensory modalities are imaginable for sensory augmentation, tactile augmentation devices have preferentially been used in many research setups (Bach-Y-Rita et al., 1969; Van Erp and Van Veen, 2003; Tsukada and Yasumura, 2004; Lindeman et al., 2005; Nagel et al., 2005). Besides academic research, the field of sensory augmentation recently also gained a lot of interest from industry. Many big companies lately introduced devices for augmented reality (e.g., Google Glasses, Microsoft Hololens, BMW Augmented Vision). However, while more products hit the market, there is a poor understanding of the underlying behavioral and neuronal mechanisms that reflect the process of combining the augmented and native senses.

Recently, Kaspar et al. (2014) performed a longitudinal study with a tactile augmentation device (the feelSpace belt) and reported that subjects developed an altered perception of space after a few weeks of training. Furthermore, it has been shown that tactile augmentation is particularly useful both in a visual search task (Wahn and König, 2016) as well as when participants are deprived of visual information (Faugloire and Lejeune, 2014). Hence, for the purpose of the current study, we built a rotating platform that was linked to a tactile augmentation device. In particular, we aimed to investigate whether people instantly combine an augmented tactile sense with vestibular information on whole body yaw rotation. Also none of the participants received any training with the augmentation device, as we intended to investigate the ability to instantaneously integrate augmented and native sensory information, rather than long-term training effects. As the main goal, we then compared prediction performance for the bimodal condition between a "winner takes all" (WTA) model

and three more complex models: The Bayesian optimal integration model, and two types of Bayesian alternation models, one using objective measured weights and the other using subjective weights obtained via a questionnaire.

Methods

Tactile Augmentation Device and Rotating Platform

Altogether we tested our participants in three conditions: rotating on the platform (native condition), receiving tactile vibrations around their waist (augmented condition), and both, rotating on the platform with simultaneous tactile vibrations (bimodal condition). Similarly to other setups in multimodal research, we employed a two-interval forced choice paradigm and tested participants in the two unimodal conditions (native or augmented) and the one bimodal condition (native plus augmented). Importantly, the tactile augmentation device and the rotating platform were precisely synchronized such that both signals provided redundant information. The tactile augmentation device (hereafter referred to as "tactile belt") can, as the name suggests, be worn around the waist. An external computer controlled all 32 vibro-motors remotely via a serial port connection. The belt itself (Figure 4.1A) is made of a flexible fabric such that people with different abdominal sizes could wear it comfortably and the angular distance between two neighboring vibro-motors remained constant (11.25°). During the whole experiment, all participants wore the belt just above their t-shirt or undershirt so that the elicited vibrations could be felt easily. When the tactile belt was switched on, at all times exactly one vibro-motor was active. For example, a rotation of 180° was accompanied by successive activation of half of the vibro-motors. Belt design and technology have been described in detail before (Nagel et al., 2005; Kärcher et al., 2012). To experimentally control angular rotation, we built a rotating platform with a chair fixed in the middle of it (Figure 4.1B). The platform could be remotely controlled, and precise parameters about angle and speed were adjusted on a trial-to-trial basis. Importantly, in the bimodal condition the vibration direction of the tactile belt was opposite to the rotation direction of the platform. In our setup, participants sat on the chair, were blindfolded, and wore headphones through which we played pink noise. Additionally, all participants held a response box with both hands, which was used for giving the

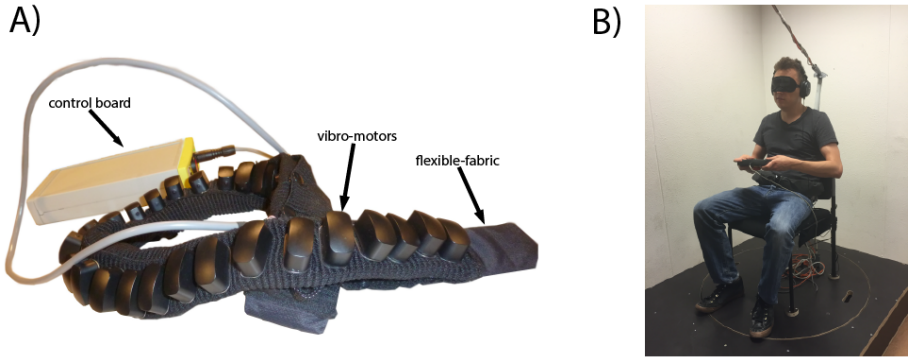


FIGURE 4.1: **Sensory augmentation device and rotating platform.** (A) Shows the tactile sensory augmentation device with its main components. (B) Illustrates the experimental setup. A participant is sitting on the chair fixed on the rotating platform and is wearing the tactile belt. He is additionally provided with an eye mask and headphones for noise cancelation. The participant is holding the response box in his hands.

required responses. They either pressed the left or the right button (indicating selection of the first or second rotation, respectively). A consecutive press on a button in the middle started the next trial.

Experimental Paradigm

Overall, the trial design was similar for all three experimental conditions. Importantly, the task and the information provided was identical, however, what varied between conditions was the type of sensory modality by which the information was provided. In the augmented (tactile-only) condition, only the belt vibration was activated. Here, participants had to judge angular differences purely based upon the successive tactile vibrations. In the native condition, only the platform rotated, so that the subjects had to rely only on rotational information. In the bimodal condition both the tactile belt and the platform were switched on synchronously and, therefore, subjects could use both sources of information. In all conditions, a trial consisted of two consecutive rotations (in the augmented condition only successive vibrations) with different angular sizes, with a one-second inter stimulus interval in between. The participants'

task was to choose either the first or the second rotation (2 IFC task) depending on which of the two rotation angles was bigger. The participants had to press the left button to indicate that the first rotation was larger or the right button to indicate that the second rotation was larger. In fact, in half of the trials the first rotation was larger, and in the other half the second one was larger. After making their choice, the participants confirmed it by pressing the center button, whereupon which the next trial started immediately. Each trial consisted of a reference and a comparison stimulus. The reference stimulus was fixed at 146.25° (equivalent to a distance of 13 intervals between the vibro-motors) and kept constant throughout the whole experiment. The comparison stimulus varied in steps of 11.25°, the distance between two adjacent vibro-motors. The order of the reference and the comparison stimulus (i.e., which of the two was the first rotation) switched randomly on a trial-to-trial basis and was balanced overall for each subject and condition. We implemented 11 different combinations of rotation angles, ranging from five steps less than to five steps greater than the reference value (90°–202.5°), plus the condition in which both reference and comparison stimuli were identical. Each of these angle combinations was repeated 10 times in a random sequence within one modality condition. For all these trials, the speed was set constant to about 42°/s and the direction of the rotation was the same within a trial, but varied (in a balanced way) across trials. Additionally, we included 10 catch trials in each condition, for which we changed the speed between the two rotations (42° vs. 32°/s). Contrary to the "normal" trials, in the catch trials the shorter rotation (in time) was associated with a wider rotation (in rotational angle) and vice versa. This was used to evaluate how much each condition was influenced by cognitive strategies (e.g., counting time). Catch trials and normal trials were randomly intermixed. Altogether, each condition consisted of 120 trials that were recorded in a block. Participants were offered a chance to take a short break after each set of 40 trials and a larger break after each block (condition). The breaks within one block ranged from about 20 s to about 2 min depending on the subjects arousal level. The breaks between blocks ranged between 1 and 5 min also depending on the subject. Each session, including all three conditions, lasted for about two and a half hours. All participants came to the lab three times, and on each visit all three conditions were measured. The order of the conditions was balanced across subjects.

Participants, Data Cleaning, and Questionnaire

Overall, 30 subjects were recorded within a period of about 5 months. However, two subjects did not complete all sessions, which left us with 28 complete data sets (16 participants were females). The age grand average of these participants was 24.03 years (SD = 3.3 years). All of the participants were students at the University of Osnabrück and each subject received either 40 euros or eight "participant hours" (which are mandatory for psychology and cognitive science students) as reimbursement for their participation. Prior to the recordings all participants were informed about the purpose of the study and signed information and consent forms. Furthermore, ethical approval was obtained by the university institutional review board. Although, we tried to make the experience with the tactile belt as comparable as possible for all participants, subjective tactile sensation was arguably rather diverse. Hence, we removed the data of five participants for which the just noticeable difference (JND) could not be determined or could be determined only with very high uncertainty. These participants presumably had difficulties processing the tactile stimulus or misunderstood the task and the inclusion of their data would thereby decrease the plausibility of consecutive analysis. This procedure ensured that later analysis was based on robust measures. This left a total of 23 participants for the remaining analysis.

In addition to the two-interval forced choice task, all participants were required to fill out a questionnaire after each condition. The questionnaires were designed to find out how intuitive and difficult each condition was and how participants judged the reliability and relevance of the provided signals. Almost all questions were defined on a Likert scale (1–5) such that participants had to choose how much they agreed with a certain statement. The questionnaire was identical for the three sessions and most of the questions were also identical between conditions. For instance: "the task was difficult," or "I was confident about my answers." A few other Likert questions varied slightly between conditions, e.g., "The belt's signal was intuitively understandable" vs. "The rotation signal was intuitively understandable," "The belt's signal was prominent in my perception" vs. "The rotation signal was prominent in my perception." Besides the Likert based questions we also asked the participants to tell us which strategy they used from a fixed set of options (the complete

questionnaire is provided in Section questionnaires in the Appendix B of Supplementary Material). Completing a questionnaire after each condition and session all subjects filed out nine questionnaires in total.

Analysis

The main analysis procedure can be summarized in three main steps: First, the JND, the Point of subjective Equality (PSE), and the uncertainty of the JND for each condition and subject were estimated using a probit model. Second, based on the observed unimodal JNDs, we calculated the predicted bimodal JND (individually for each subject) for all tested models. Third, using observed and predicted bimodal JNDs, we calculated the reduced chi-squared statistic (χ^2_{red}) for each model. The next two paragraphs will explain these steps in more detail. Furthermore, we describe the questionnaire analysis in Section Questionnaire Analysis.

Curve Fitting In order to calculate the JND, we fitted for each subject and condition a GLM with a probit link to the behavioral data. The function is formalized in Equation 4.1, where β_1 and β_2 are the two (optimized) parameters of the model fit and the "norminv" function computes the inverse of the normal cumulative distribution function (cdf). We needed to invert Equation 4.1 in order to obtain the corresponding value of angular difference (x) for a specific performance level. Then, we used the asymptotic threshold of one standard deviation of a cumulative binominal distribution function (84%) as the corresponding angular difference of Equation 4.2 (y) and, consequently, calculated the JND. This gave us a direct measure of how precise each subject was able to distinguish the two angular stimuli from each other, separately for each condition. Next, we estimated the quality of the estimate of JNDs. Hence, we applied the error propagation method using the matrix formalism as described by Equation 4.3. Here, U_{JND} represents the uncertainty of the JND estimation, V_β is the covariance matrix of the betas, and $J_{i,j}(\beta)$, shown in Equation 4.4, is the Jacobian matrix. As an example plot Figure 4.2 demonstrates that

most of the participants showed a behavior well-described by typical sigmoidal psychometric function.

$$\beta_1 + x \cdot \beta_2 = \text{norminv}(y) \tag{4.1}$$

$$x_{JND}(\beta_2) = \frac{\text{norminv}(y_{\text{theshold}})}{\beta_2} \tag{4.2}$$

$$U_{JND} = \sqrt{\text{diag}(J(\vec{\beta})V_\beta \cdot J^T(\vec{\beta}))} \tag{4.3}$$

$$J_{i,j}(\beta_1, \beta_2) = \frac{\partial x_{JND}}{\partial \beta_j}(\beta_1, \beta_2), j \in \{1, 2\} \tag{4.4}$$

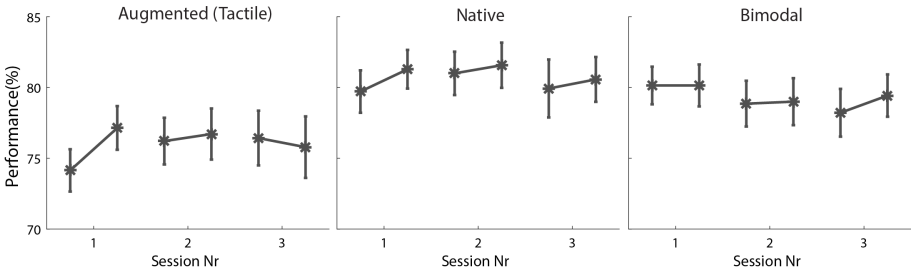


FIGURE 4.2: **Example logistic fit (native condition).** The figure demonstrates the performance of one participant in the native condition as an example. The abscissa illustrates the difference between the two angular rotations (reference—comparison angle) in degrees. The ordinate indicates the probability to choose the reference angle. The green circles show the recorded behavioral data, the solid red curve shows the logistic fit, and the dashed red lines indicate the uncertainty of the fit. The magenta line depicts the Point of subjective Equality, while the blue line depicts the sensory threshold, at one standard deviation (84%) of the psychometric function. The distance between the PSE and the intersection of the blue line with the abscissa represents the JND.

Model Comparison The different models varied in their mathematical complexity for predicting bimodal performance. The simplest model was a static/intercept model which predicted always the same (mean) value for all subjects. The next one was a WTA model, which took either the native or the augmented JND (depending which of them was smaller) to predict bimodal JND. The Bayesian optimal integration model and the Bayesian alternation models were more complex. The Bayesian optimal integration model can be expressed as shown in Formula (5), while the Bayesian alternation model is described in Formula (6). μ illustrates the PSE of the psychometric function while P stands for the probability (i.e., relative weight) of each modality. In both Formulas 4.5 and 4.6, σ represents the JND, na is the abbreviation for the bimodal (native plus augmented) condition, n equals the native-only condition, and a stands for the augmented-only condition.

$$\sigma_{na}^2 = \frac{\sigma_n^2 \cdot \sigma_a^2}{\sigma_n^2 + \sigma_a^2} \tag{4.5}$$

$$\sigma_{na}^2 = P_n(\mu_n^2 + \sigma_n^2) + P_a(\mu_a^2 + \sigma_a^2) - (P_n \cdot \mu_n + P_a \cdot \mu_a)^2 \tag{4.6}$$

An interesting question regarding the Bayesian alternation model is how to determine the probabilities for the two unimodal modalities, Pn and Pa. We decided to implement two different approaches. On the one hand, we used the observed objective (although subject specific) reliabilities such that the native probability could be formulated as described in Equation 4.7 and, analogously, the augmented probability as in Equation 4.8.

$$P_n = \frac{\frac{1}{\sigma_n^2}}{\frac{1}{\sigma_n^2} + \frac{1}{\sigma_a^2}} \tag{4.7}$$

$$P_n = \frac{\frac{1}{\sigma_a^2}}{\frac{1}{\sigma_a^2} + \frac{1}{\sigma_n^2}} \tag{4.8}$$

On the other hand, we calculated native and augmented weights on the basis of the individual questionnaire responses. For this procedure, we selected

the following eight performance relevant questions of the native and the augmented parts of the questionnaire ("I have done similar tasks before," "The task was intuitive," "The task was difficult," "I think I performed well in the task," "I was confident about my answers," "I felt comfortable with the task," "The belt / the rotation gave me relevant information to solve the task," "The belt/the rotation signal was prominent in my perception"). As all these questions were answered on a Likert scale, we could directly apply mathematical operations on them. First, we averaged the responses of the three different sessions separately for each question and then subtracted answers relating to the native condition from those relating to the augmented condition. As a result, for each question we knew whether the augmented or the native task was more intuitive, difficult, and so on (positive numbers indicated higher agreement in the augmented task, negative numbers indicated higher agreement in the native task). In order to combine the responses of all questions, we applied a principal component analysis resulting in eight different components. To further reduce dimensionality and to calculate subjective weights we then considered only the first component for further processing. Through this procedure we reduced all questionnaire responses to one scalar per subject. Finally, we normalized this number to the range of zero to one using a logistic function. These values were then used as augmented weights P_a. The native weights P_n were then defined as the inverse $1 - P_a$. Although the complexity in terms of the mathematical expression varied between the models, we want to emphasize that we did not optimize free parameters for any model. In summary, we optimized the estimation for the observed JND, but we did not fit/improve unimodal to bimodal prediction performance by adjusting model parameters. Hence the amount of free parameters (k) was zero, and consequently the degrees of freedom ($v = 22 = N - k - 1$) were constant throughout all investigated models.

Questionnaire Analysis The main goal of the questionnaire analysis was to create a deeper understanding of the quantitative measurements. Therefore, we first looked at single questions in the unimodal parts and examined possible differences between the augmented and native ratings. Second, we analyzed the categorical responses about strategy use in all conditions in order to get a better estimate of how each participant subjectively approached the task. Here, all subjects had to choose one out of the following options: (a) tactile cue only, (b) rotation only, (c) combination of both cues, (d) counting time, (e)

visualization, (f) random guessing, (g) other. We decided to focus only on these two analyses in order to keep a clear structure.

Results

Control Statistics

First of all, we aimed to investigate whether the subjects exhibited non-stationarities within and/or between sessions, for example, in the form of learning or fatigue effects. Hence, we split the data for each condition and session into the first and second "block-half" and performed three separate repeated measures analyses of variance (one for each condition) with session and block-half as independent (repeated) variables and the number of correct responses as dependent variable. Catch trials were not considered in this analysis. However, as shown in Figure 4.3, neither session nor block-half nor the interaction of both factors revealed a significant influence in any of the three conditions (please find the analysis of variance tables in the Appendix B in Supplementary Material). This indicates that the subjects' performances were constant within and between sessions. As the data did not reveal any indication of learning or fatigue effects, we then collapsed the data over all three sessions and calculated the amount of correct responses separately for each angle combination.

Comparing Conditions

After calculating the JNDs and PSEs (see Section Curve Fitting), we compared both measures between experimental conditions. Figure 4.4A displays the results for the PSE. First we applied separate t-tests for each condition to test whether the PSE was different from zero. While for the augmented condition $[t(22) = 0.109, p = 0.914]$ and the bimodal condition $[t(22) = -1.726, p = 0.098]$ the PSE was not significantly different from zero, the native condition revealed a significant difference $[t(22) = -7.422, p \leq 0.001]$. Furthermore, we analyzed the PSE using a repeated measures ANOVA for the factor condition, which revealed a significant effect $[F(2, 44) = 7.976, p = 0.001$, partial $\eta^2 = 0.266]$. Post-hoc comparisons confirmed a significant difference between the native and the augmented PSE ($p = 0.001$), but no significant

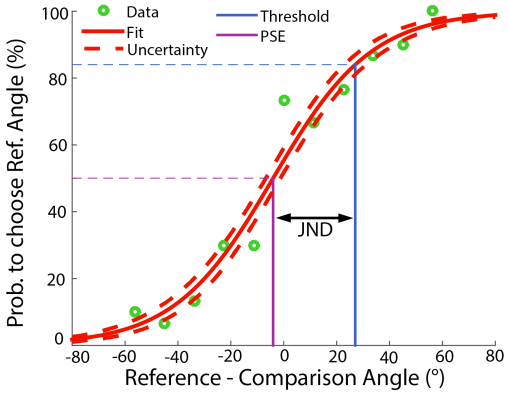


FIGURE 4.3: **Investigation of learning effects.** The abscissa divides the data of the three different sessions and the data of each session between the first half and second half of the block, separated by condition. The ordinate indicates the performance as a percentage. The error bars illustrate the average performance with the error bars representing the standard error of the mean.

difference between the augmented vs. bimodal PSE ($p = 0.101$). The native vs. bimodal contrast was borderline non-significant ($p = 0.051$). Importantly, as a measure of the subjects' performance we analyzed the JND; Figure 4.4B illustrates these results. For the statistical analysis of the JND we also applied a repeated measure ANOVA. The result revealed a main effect of condition [$F(2, 44) = 17.869, p < 0.001$, partial $\eta^2 = 0.448$]. Post-hoc pair-wise comparisons confirmed that the JND in the augmented condition was higher than in the native ($p < 0.001$) and the bimodal ($p = 0.010$) conditions. The JND in the bimodal condition was, in turn, higher than in the native condition ($p = 0.003$). Hence, the native condition resulted in the best performance, followed by the bimodal condition; the augmented (tactile) performance was the worst. This rather compelling result indicates that native and augmented sensory modalities were not combined in a "Bayesian optimal way," as this would require that the bimodal JND is less or equal than in either single modality. This raises the question of alternative models to be compared in the following investigation. As it is the gold standard in many studies on multisensory integration, we kept the Bayesian integration model in the model comparison procedure and

compared it to several alternatives as described in the next paragraph.

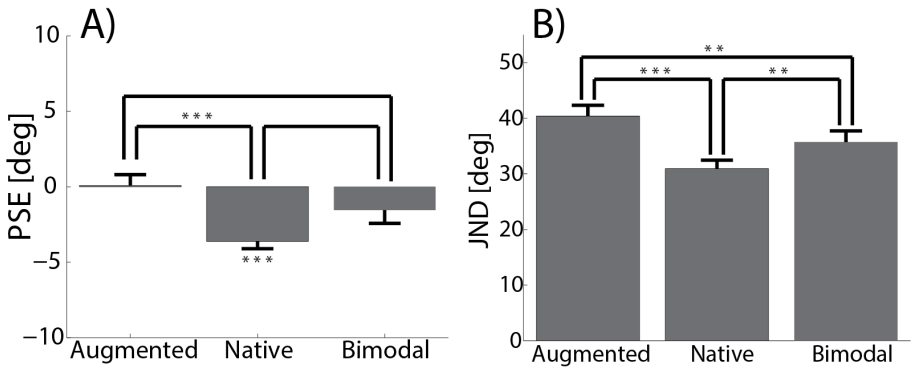


FIGURE 4.4: **Comparing conditions.** (A) Shows the PSE (on the ordinate) separately for the three different conditions on the abscissa as a mean over subjects. The asterisks below indicate the significance level for the difference of the SPE to zero. The asterisks above show the significance level for the comparisons between conditions. (B) Shows the JND again separately for the three different conditions on the abscissa and as a mean over subjects. The asterisks illustrate the level of significant differences between conditions.

Model Comparison

The main goal in our study was to determine the cognitive mechanism that underlies the combination of the augmented and native sensory cues provided. To address this central question of the study, we compared the five different models in their accuracy to predict the bimodal JND given the unimodal JNDs [Intercept, Winner Take All (WTA) optimal Bayesian integration, objective Bayesian alternation, subjective Bayesian alternation]. In particular, we combined the model prediction with the uncertainty measurement to calculate the reduced chi-squared value (χ^2_{red}), as shown in Formula 4.9.

$$\chi^2_{red} = \frac{1}{V}\sum_{k=1}^{n}\frac{(\text{JND}^2_{\text{observed}} - \text{JND}^2_{\text{estimated}})^2}{(\text{JND}^2_{\text{uncertainty}})^2} \tag{4.9}$$

This gave us a measure of how much variance each model could explain compared to the optimum (χ^2_{red} = 1), when all structure is explained and the residual variance is due to noise only. Our results show that the intercept model is a poor fit for the data (χ^2_{red} = 10.95) and leaves a lot of variance to be explained. Figure 4.5 summarizes the result for the other four models of interest. Although the Bayesian integration model (Figure 4.5A) is clearly a better model than the intercept model, it also leaves quite some variance to be explained (χ^2_{red} = 4.34). While the objective Bayesian alternation model outperformed the integration model (Figure 4.5C, χ^2_{red} = 1.67), the WTA model predicted bimodal behavior even slightly better (Figure 4.5B, χ^2_{red} = 1.64). However, using subjective weights for the Bayesian alternation model, prediction performance could be significantly improved such that it had the highest prediction rate and lowest residual variance among all tested models (Figure 4.5D, χ^2_{red} = 1.09). In fact, the χ^2_{red} of the subjective Bayesian alternation model is very close to the optimum of χ^2_{red} = 1.00.

Subjective vs. Measured Reliabilities

To better understand the differences between the two types of Bayesian alternation models, using different unimodal weights Pn and Pa, we implemented an optimization procedure to find the weights that yielded the optimal prediction accuracy for each subject (referred to as the "optimal predicted weights"). That is, we did not investigate how subjects could perform optimally, but which type of weights for native and augmented modality (per subject) would optimally explain the data as they were observed. The two weights that were used in the model comparison procedure (objective and subjective) were then analyzed against these optimal predicted weights using a linear regression. As shown in Figure 4.6A, the comparison of the optimal predicted weights and the objective (based on unimodal performance) weights were uncorrelated [r(22) = 0.0008, p = 0.887]. In constrast, the subjective (questionnaire-based) weights showed a strong and significant correlation with the optimal predicted weights [r(22) = 0.4782, p < 0.001, Figure 6B]. This result indicates that compared to the objective/measured reliabilities, the subjective evaluations (weights) better captured the intersubject variability of cue preferences.

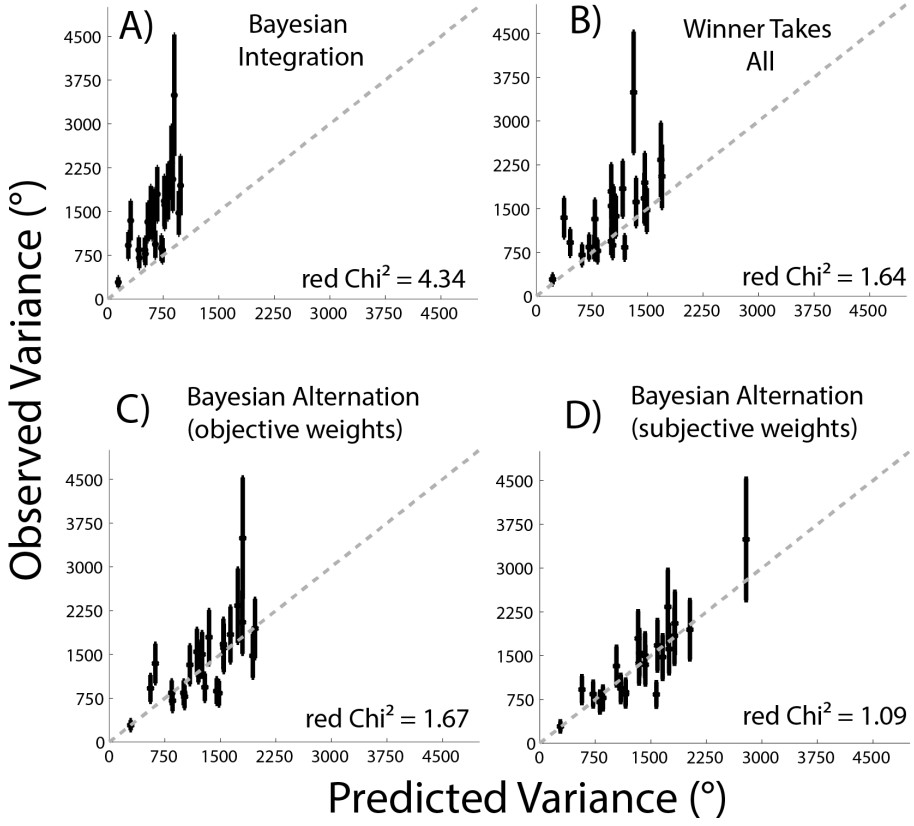


FIGURE 4.5: **Model comparison.** The abscissa shows the predicted squared JND; the ordinate shows the observed squared JND in the bimodal condition. Each black dot shows the predicted vs. the observed value for one subject. The error bars around the black dots illustrate the uncertainty of the observed bimodal values. The gray dashed diagonal line represents the ideal prediction. The resulting χ^2_{red} is plotted for each model. (A) Bayesian integration. (B) Winner takes all. (C) Bayesian alteration (objective weights). (D) Bayesian alteration (subjective weights).

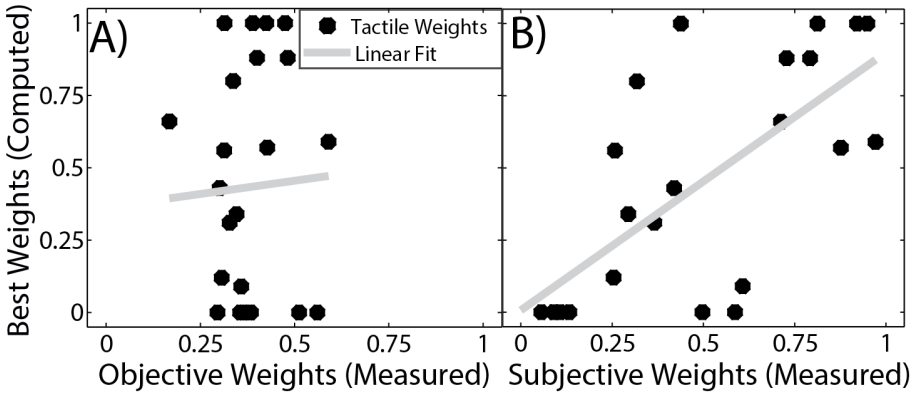


FIGURE 4.6: **Correlation of measured and computed weights.** The ordinate shows the individual tactile weights that lead to the best possible model fit of the Bayesian alternation model. (A) Shows the correlation of the best weights and the actual measured weights based on measured reliability. (B) Shows the correlation of these estimated weights and the weights calculated from the questionnaire data. Each black dot represents one participant. The gray lines show the least square linear fit to the data.

Strategy Assessment

All participants were deprived of visual information and could not use auditory information due to the pink noise played on the earphones. As angular rotation activates the semicircular canals, which are part of the vestibular system, the main sensory input here was the vestibular modality (for that reason we use the term native modality). However, processing angular rotation without visual information might have been a rather unusual experience for many of our participants. In theory, participants could therefore have also used some more cognitive strategies like counting time or visualizing images. In order to address this question, we analyzed the catch trials and the subjective questionnaire data (directly asking for the strategy employed). For the catch trial analysis, as shown in Figure 4.7A, we compared the performance in three types of trials: first, the performance in the catch trials itself, which had reversed angular-time differences; second, the performance in trials with the same angular difference as in the catch trials (11.25°) but a much shorter time difference (250 ms);

and third, the performance in trials with 45° angular difference, as they were most similar in the time domain to the catch trials (1100 ms difference), but very different in the angle domain. Figure 4.7B shows that in the augmented tactile task, performance in the catch trials (blue) was more similar to the trials with the same angle difference (green), compared to the trials with the same time difference (red). This supports the view that subjects used angular but not time information in this condition. However, this pattern was reversed in the native task, such that catch trial performance in the native task was more similar to same-time trials. Hence, the native task was clearly influenced by time (counting) information. The bimodal task performance was again in between these two, with a trend toward the angle-based trials, supporting the idea that signal/strategy usage alternated on a trial-to-trial basis. Figure 4.7C illustrates the results of the subjective strategy assesment. In the respective (bimodal) question "Which strategy did you use to solve the task?" all subjects had to choose one out of following the options: belt only, rotation only, combination of belt and rotation, time counting, visual imagination, random guessing, other strategy. As the last three options (visual imagination, random guessing, and other strategy) were chosen only rarely (each < 8%) we summarized them to "other strategies." Overall, the results of the questionnaire analysis are in line with the catch trial analysis. Subjects reported to have used the time information in only 8 out of 69 sessions for the augmented tactile task. In the bimodal condition, subjects reported that counting time was their preferred strategy in 12 out of 69 sessions. Again, the native task showed a reversed picture. Here, participants reported that they were counting time in 38 out of 69 sessions (23 subjects * 3 sessions = 69 total sessions). Overall, both questionnaire and catch results indicate that most subjects relied on cue processing in the augmented and bimodal tasks and suggest that counting time and other cognitive strategies played a major role in the native condition.

Relevance and Dominance of the Signals

As a final step, we investigated differences between augmented and tactile questionnaires for individual questions. Most questions did not reveal interesting or significant differences between the augmented condition and the native condition. Figure 4.8 contrasts the agreement (mean over subjects) for the following four questions between the native and the augmented condition: 1.

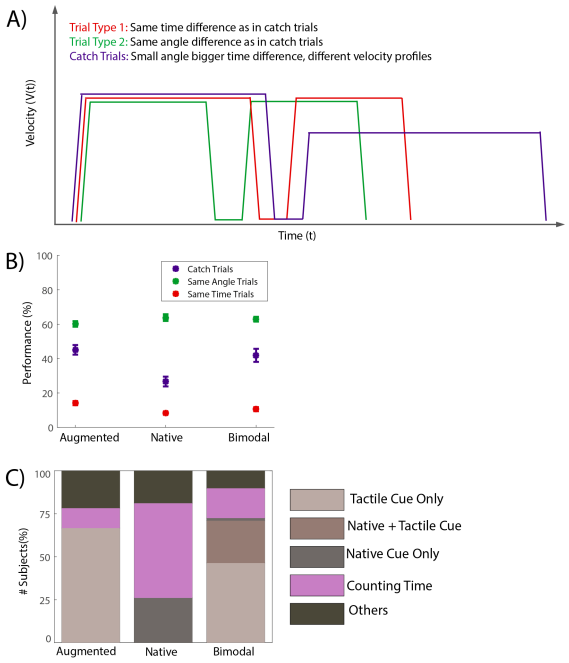


FIGURE 4.7: **Strategy assessment.** (A) Displays the velocity profiles of the different trial types that were analyzed in the catch trial analysis. The green line represents a "hard trial," with small angular difference. The red line shows an easy trial with large angle (and time difference). The blue line illustrates the catch trial with inverted angle-time difference (the shorter rotation took more time). (B) Displays the result of the catch trial analysis. The abscissa separates the three experimental conditions. The ordinate illustrates average performance across subjects. The blue error bars show the performance in the catch trials, the red error bar shows the performance for the easy trials (same time difference as catch trials), and the green error bars show the performance for the hard trials (same angular difference as the catch trials). (C) Shows the result of the questionnaire analysis regarding subjective strategy use. The abscissa again separates the three experimental conditions and the ordinate indicates the proportion of subjects using a particular self-assessed strategy. The different strategy types are color coded and labeled.

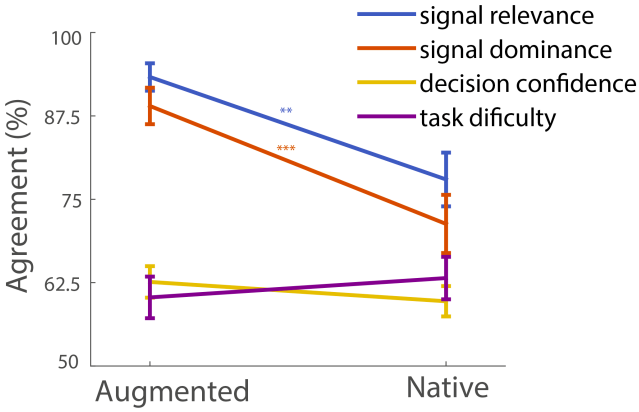


FIGURE 4.8: **Signal perception.** The abscissa separates the tactile and the vestibular condition. The ordinate indicates the level of agreement for a certain question. The error bars are standard errors of the mean. The asterisks indicate the level of significant difference between the two conditions.

"The belt/the rotation gave me relevant information to solve the task," 2. "The belt/the rotation signal was prominent in my perception," 3. "The task was intuitive," 4. "The task was difficult." The analysis showed that the native task was perceived as more difficult; however, the difference from the augmented task was not significant [t(22) = -0.91, p = 0.373]. In line with this observation, subjective decision confidence was higher in the augmented condition, but again failed to reach significance [t(22) = 1.21, p = 0.2402]. However, two other questions showed clear effects; the first was signal relevance, the other was probing signal dominance. Participants judged the tactile belt as providing information with higher task relevance compared to the angular rotation of the platform [t(22) = 3.34, p = 0.0030]. Similarly, the belt was rated to be perceptually more dominating [t(22) = 4.36, p = 0.0002] than the native information.

Discussion

Summary

We tested whether untrained adult participants are able to use augmented tactile information in a two-interval forced choice task and examined how such augmented information is combined with information from native senses. Psychometric data and consecutive statistical analysis show that all subjects were able to solve the task using only the tactile information from the augmented sense. Hence, even without prior information or experience the participants were able to use the supplied augmented tactile information for the current task. The model comparison demonstrated that the subjective Bayesian alternation model had the highest prediction performance. This model reflects the idea that on each trial a subject is using one or the other sensory signal provided, caused by a (probability based) Bayesian selection mechanism. This finding is in line with earlier findings on Bayesian Alternation observed in children (Gori et al., 2008; Nardini et al., 2008). However, a more precise look at the data revealed that about half of the subjects strongly preferred one of the cues (native or augmented) while the other half used both cues more evenly. As a result, the respective weights for subjects with such strong preferences are matching a simple winner take all (WTA) strategy where these weights are set to one and zero, respectively. Although such behavior could be described with a much simpler WTA model, the Bayesian alternation model yields clearly higher prediction performance on a group level. This was due to the fact that the other half of the subjects alternated between both cues more often so that this behavior was better captured with the Bayesian alternation model. Altogether one can say that the spectrum of cue preferences was rather continuous between subjects. Some subjects preferred the augmented cue, some others preferred the native cue, and again many others were in-between these extremes. To put it differently, one can consider the subjective Bayesian alternation model as an extension of a WTA strategy. In particular it is more flexible as it allows to alternate signal usage on each trial compared to each subject (but doesn't require it). Moreover, we would like to point out that this does not involve fitting free parameters, but is purely based on observed unimodal performance and questionnaire data. The difference in prediction performance between objective and subjective alternation model is of further

interest. Hereby, we demonstrated that a combination of qualitative and quantitative data represents an interesting and advantageous method in the field of multimodal research.

Potential Shortcomings of the Study

One concern in the current study is related to the native modality condition. Although sensory input was provided only to the vestibular system, roughly half of the participants presumably involved a cognitive strategy such as counting time. However, the vestibular system necessarily has to integrate information over time and cannot provide an absolute reference. Hence, it would not be reasonable to assume that time information does not play any role for the vestibular system. From this perspective, we argue that the vestibular sense is to some extent a "time-angle integrator." This idea was also discussed in a study by Berthoz et al. (1995). Furthermore, the catch trials were the hardest trials to solve with an angle-based strategy, as they not only had a very small angular difference, but also particularly long rotations (202.25° vs. 191°). Hence, subjects who aimed to use angular information in catch trials basically had to guess. As a result, some subjects might have used the counting strategy mostly in the catch trials in order to avoid guessing. Grondin and colleagues showed that humans benefit from a time counting strategy especially when judging intervals longer than 2.5 s (Grondin, Meilleur-Wells, and Lachance, 1999). Similarly, Clément and Droit-Volet showed that adults temporal sensitivity increases with explicit time counting, while this is not the case for children (Clément and Droit-Volet, 2006). In fact, fewer participants subjectively claimed in the questionnaire to have used a time-counting strategy as the catch trial analysis suggested. In conclusion, we argue that the native condition was influenced by both vestibular signals as well as higher cognitive strategies, in particular counting time. However, cognitive strategies did not make a major contribution in the augmented (tactile) or in the bimodal condition. Altogether, the investigation of signal/strategy use supports the idea that the majority of subjects used a subjective Bayesian alternation process to combine both sensory stimuli.

A second issue concerns the tactile belt and differences among individuals. Although, we tested the augmentation device before each session, some subjects reported that they sometimes did not properly feel the vibration. Differences

in waist size, position of the vibro-motors, undershirt material, and the participant's ability to differentiate tactile stimulations might have altered perception of the tactile sensation. Due to technical limitations, in some cases one or the other vibro-motor might also have vibrated less strongly than others. To counteract these issues, we removed participants for whom we could not reliably estimate the psychometric performance (JND) that was later needed in the model comparison and other analysis (as described in the Method Section). Our results clearly show that most subjects follow a subjective Bayesian alternation strategy for the combination of native and augmented sensory cues. However, differences among individuals were strong enough that our conclusion is reasonable for the majority of the subjects, but not for each and every individual. Weights that led to an optimal prediction showed that many participants strongly preferred the augmented cue, while other subjects had a clear bias in favor of the native signal, and still others lay in between these two extremes. Importantly, the subjective questionnaire data helped to better understand those individual differences in performance measures and time-counting strategies. All in all, subjective and objective measurements nicely match and complement each other and hereby create a more complete picture of the reported findings.

Integration vs. Alternation

Many studies suggest that human multimodal processing involves "optimal integration." Without arguing against such an overwhelming and high quality amount of empirical evidence, our findings qualify this statement to some extent. In fact, a closer look into the literature reveals that several studies reported deviations from the "standard" Bayesian integration model. One of best examples were reported by Nardini and colleagues as well as Gori and colleagues, both providing clear evidence that optimal integration is not present in children until the age of 8 years (Gori et al., 2008; Nardini et al., 2008). Most recently, Adams compared different integration models with an audio-visual temporal judgment task and similarly reported that older participants employed a partial integration strategy while younger participants (<8 years) did not integrate, but instead switched between the two sensory signals provided (Adams, 2016). Besides research with infants, there is evidence that under certain circumstances even adults do not integrate, but instead alternate

between two sensory cues. In particular, de Winkel et al. (2013; 2015) performed a visual-vestibular cue combination task in which adult participants were rotated around the yaw axis, given either additional visual information or not. Most interestingly, the authors reported that only about half of the participants behaved in congruence with the Bayesian integration model, while the others most likely alternated in the usage between the two cues (De Winkel et al., 2013; De Winkel, Katliar, and Bülthoff, 2015). One of the possible explanations for both, our results as well the experiments from de Winkel et al. (2013; 2015) would be that the two sensory signals were not perceived to have a common cause (Körding et al., 2007), although they were supplying redundant information. As Ernst (2007) and Kaliuzhna et al. (2015) showed, it is possible that humans (directly) integrate two arbitrary associated sensory signals. However, combining rotational information and (augmented) tactile stimulation might require a more complex mapping than the visual-haptic associations used in these studies. An interesting idea for a follow up study of our paradigm might be to explicitly force the integration, or at least comparison of both cues. In such a scenario the information in the first interval could be provided tactilely, while for the second interval the information would be displayed via the platforms rotation (or vice versa). How well participants can solve such a task needs to be addressed in future research.

Multisensory Learning

The comparison of prediction performance between the Bayesian integration model and the Bayesian alternation model showed that participants in our study most likely alternated between using augmented and native information. Research with infants has provided evidence that optimal integration of sensory cues is not a native mechanism, but instead has to be acquired (Gori et al., 2008; Nardini et al., 2008). Moving to the other side of the age spectrum, Bates and Wolbers recently showed that the combination of visual and self-motion cues becomes less than optimal with age. The authors attribute this observation to neural degeneration in entorhinal and hippocampal regions (Bates and Wolbers, 2014). Accordingly, neural degeneration and atrophy were shown to increase with age (Dickerson et al., 2001). In general, recent studies have shown that multisensory influences arise relatively early and by a variety of mechanisms (Driver and Noesselt, 2008). In a review from 2008, Stein and

Stanford argued that many multisensory neurons exist in the superior colliculus. Explicitly they showed that this region combines visual, auditory, and somato-sensory input to control eye and head movements (Stein and Stanford, 2008). Burnett and colleagues tested this assumption by lesioning cats superior colliculus and conclude that damage to this area directly causes a loss of multisensory neurons which again led to a decrease of multisensory behavior (Burnett et al., 2004; Burnett et al., 2007). Hence, one can conclude that optimal cue integration is experience dependent and relies on intact neural structures.

While children presumably take a couple of years to successfully integrate information originating from two native modalities, it has been unclear until now how such a process is established with an augmented sense in adults. Here, we provide the first evidence that the majority of adult participants combine augmented and innate sensory modalities using a subjective Bayesian alternation strategy. However, we speculate that intensive training with the sensory augmentation device could lead to a shift in the cue combination strategy. Specifically, over time Bayesian cue alternation might be replaced by optimal Bayesian cue integration, which might be associated to casual inference mechanisms described by Körding et al. (2007). In such a scenario the augmented tactile stimulation would improve overall performance. In line with this idea, several studies showed that training alters the individual reliabilities in a cue combination paradigm (Jacobs and Fine, 1999; Atkins et al., 2001). Furthermore, Shams and Seitz (2008) provided striking evidence that multimodal learning is more effective than unimodal learning. Hence, as a next step we plan to conduct a longitudinal study and investigate how training with the augmentation device will change cue combination strategies in adults.

Cue Combination and Attention

There has been a long debate whether attentional resources share a common reservoir (Jolicoeur, 1999; Arnell and Larson, 2002) or whether each modality has its own attentional resources (Potter et al., 1998; Talsma, Doty, and Woldorff, 2007; Martens, Kandula, and Duncan, 2010; Wahn and König, 2015b; Wahn and König, 2015a; Wahn and König, 2016). In fact, attention might have played an important role also in our study. Subjects reported that

the tactile stimulation dominated their perception to a significantly stronger degree than did the angular rotation. Hence the participants' attention was driven toward the tactile stimulation. The observed Bayesian alternation process can therefore also be understood as an attentional mechanism. In this view, both cues rivaled for attentional focus such that it switched on a trial-to-trial basis, with a probability that was based on subjective reliability. In conclusion, our results support the idea of a shared reservoir of attention for native and augmented sensory cues.

A second issue is concerned with the attentional load. Several studies suggested that attentional or perceptual load modulates multisensory integration (Alsius et al., 2005; Mozolic et al., 2008; Klemen, Büchel, and Rose, 2009). Oppositely, Helbig and Ernst (2008) demonstrated that haptic cue weighting is independent of modality-specific attention. Similarly, Wahn and König (2015; 2015) showed the existence of optimal integration between visuotactile and audiotactile cues even under high attentional load. In our study, attentional load was not modulated; however, considering the fact that cognitive strategies such as counting time played a major role only for the native but not for the augmented condition, future investigations with varying attentional load might reveal interesting new insights.

Subjective vs. Objective Measurements of Reliability

Our results demonstrate that objective measured reliability was higher in the native condition compared to the augmented condition. Similarly, Fetsch and colleagues demonstrated that vestibular cues are overweighted in low-reliability conditions (Fetsch et al., 2009; Fetsch et al., 2012). However, our participants reported (subjectively) that the tactile belt provided the more relevant information for the task, and the confidence ratings were slightly higher in the tactile condition. This discrepancy between the subjective awareness of a signal's reliability and its objective reliability based on the performance measurement is surprising. A direct conclusion from such an observation is that participants in our study arguably did not represent an "objective ideal observer," which many studies have proposed as a general mechanism of sensory cue combination (Blake, Bülthoff, and Sheinberg, 1993; Ernst and Bülthoff, 2004; Landy, Banks, and Knill, 2011). Opposed to that, Knill and Saunders (2003) introduced the concept of a "subjective ideal observer" who behaves optimally

according to subjective certainty. To test such an assumption, we analyzed the subjective strategy use during the presence of both signals (bimodal condition). Interestingly, most subjects claimed to have used only the belt's signal. In that sense, subjects did not behave optimally with respect to external measurements of reliability, but indeed behaved optimally with respect to the internal subjective rating of the signal's reliability. The signal that was rated to be more relevant in the unimodal conditions was used with a higher probability in the bimodal condition.

Nevertheless, the question remains as to why subjective and objective reliability measurements differ in the first place and why the subjective reliability led to increased behavioral prediction accuracy for the bimodal task. One idea would be to look at how easily and precisely the reliabilities of the two modalities can be estimated. In particular, we assume that it is advantageous to use information which is less reliable compared to information with unknown or almost-unknown reliability (no prior), even though in the end the latter might turn out to have been more reliable. In this respect, a signal's reliability might be positively biased if it can be estimated easily and quickly. On the other hand, if a signal's reliability is difficult or time-consuming to estimate (e.g., due to the lack of feedback), it might be underestimated. We argue that the belt's reliability was relatively easy to estimate for the subjects as it provides information in an absolute coordinate system and it dominated perception according to the subjective reports (as opposed to the native condition). In contrast, the reliability of the rotation information might have been quite difficult to estimate, as the vestibular system needs to integrate information over time without an absolute reference point (Barnett-Cowan and Harris, 2009). As a result, participants might have overestimated the belt's reliability and underestimated reliability based on rotation information. If such a hypothesis holds, we believe that it can have significant consequences for research investigating cue combination mechanisms and multisensory processes.

Author Contributions

CG and SP designed the experiment, recorded the participants, and analyzed the data. CG wrote the main part of the manuscript, but was supported by HF in conducting and describing the model comparisons and other mathematical

procedures. HF was also responsible for the technical setup and programmed the necessary code for both the belt and the rotating platform. PK supervised the study, corrected and edited the manuscript, and suggested the way the data should be analyzed.

Conflict of Interest Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgments

We gratefully acknowledge support by ERC-2010-AdG #269716—MULTISENSE and Cognition and Neuroergonomics/Collaborative Technology Alliance #W911NF-10-2-0022. Furthermore, we would like to thank the team of the Electronics Workshop of the Department of Physics of the University of Osnabrück. Together with the Fine Mechanics Workshop they built the rotating platform and the necessary electronic control elements. Here, special thanks go to Mr. Svajda and Mr. Lemme, who led the construction and the programming of the platform control, respectively. Finally, we would like to thank all members of the Neurobiopsychology Group at the University of Osnabrück, who helped in setting up the system and bringing in new ideas on project report meetings.

4.2 Neuronal Oscillations Indicate Sleep-Dependent Changes in the Cortical Memory Trace

This section was published as a peer-reviewed article: Moritz Köster, Holger Finger, Maren Kater, Christoph Schenk, and Thomas Gruber (2016). "Neuronal Oscillations Indicate Sleep-dependent Changes in the Cortical Memory Trace". In: *Journal of Cognitive Neuroscience*

Abstract

Sleep promotes the consolidation of newly acquired associative memories. Here we used neuronal oscillations in the human EEG to investigate sleep-dependent changes in the cortical memory trace. The retrieval activity for object–color associations was assessed immediately after encoding and after 3 hr of sleep or wakefulness. Sleep had beneficial effects on memory performance and led to reduced event-related theta and gamma power during the retrieval of associative memories. Furthermore, event-related alpha suppression was attenuated in the wake group for memorized and novel stimuli. There were no sleep-dependent changes in retrieval activity for missed items or items retrieved without color. Thus, the sleep-dependent reduction in theta and gamma oscillations was specific for the retrieval of associative memories. In line with theoretical accounts on sleep-dependent memory consolidation, decreased theta may indicate reduced mediotemporal activity because of a transfer of information into neocortical networks during sleep, whereas reduced parietal gamma may reflect effects of synaptic downscaling. Changes in alpha suppression in the wake group possibly index reduced attentional resources that may also contribute to a lower memory performance in this group. These findings indicate that the consolidation of associative memories during sleep is associated with profound changes in the cortical memory trace and relies on multiple neuronal processes working in concert.

Introduction

Sleep plays a pivotal role in the consolidation of newly acquired memories, transforming initially labile memories into more stable and lasting memory

representations (Rasch and Born, 2013; Diekelmann and Born, 2010; Stickgold, 2005). In particular, it has been demonstrated that the consolidation of explicitly encoded associative memories is promoted by slow wave sleep (Rasch, Büchel, Gais, and Born, 2007; Marshall, Helgadóttir, Mölle, and Born, 2006). Although theoretical models presume profound sleep-dependent changes in the cortical engram (e.g., Born and Wilhelm, 2012; Marshall and Born, 2007; Tononi and Cirelli, 2006), it has rarely been investigated how episodes of sleep effect memory representations in the human cerebral cortex.

According to an active system consolidation account, sleep promotes the transfer of novel memories from neuronal networks in the mediotemporal lobe into lasting memory representations in neocortical networks (Born and Wilhelm, 2012). It is assumed that the transfer of information relies on an active replay of newly encoded information during sleep, as suggested by studies in rodents (Ji and Wilson, 2007; Wilson and McNaughton, 1994) and recent evidence in humans (Deuker et al., 2013). Second, according to the synaptic homeostasis hypothesis, wakefulness is assumed to lead to a general increase in neuronal activity, whereas sleep promotes a global downscaling of synaptic strength (Tononi and Cirelli, 2006). By a proportional reduction in synaptic strength, sleep may lead to a relative enhancement of recently activated synapses. So far, this theoretical idea was investigated in animal models (Liu, Faraguna, Cirelli, Tononi, and Gao, 2010; Vyazovskiy, Cirelli, Pfister-Genskow, Faraguna, and Tononi, 2008) and recently also in the human brain in wake EEG recordings (Kuhn et al., 2016; Landsness et al., 2011). For example, Kuhn and colleagues (2016) found deceased levels of occipital theta and gamma oscillatory power after sleep, compared with sleep-deprived control group. Importantly, these two theoretical ideas are not mutually exclusive, and the consolidation of novel memories may rely on both processes working in concert (Diekelmann and Born, 2010).

A first study on sleep-dependent changes in the cortical memory representation by Gais and colleagues (2007) compared the retrieval activity before (pre) and after (post) an interval of sleep or sleep deprivation. Using functional neuroimaging (fMRI) the authors found a redistribution of retrieval activity from mediotemporal to neocortical neuronal networks, which is in support of an active system consolidation account.

Neuronal oscillations play a pivotal role in memory processes (Fell and Axmacher, 2011). In the human MEG and EEG, a particular focus lies on oscillations in the theta (4–8 Hz), alpha (8–14 Hz). and gamma (30–100 Hz) ranges. Increases in theta oscillations are associated with episodic encoding (Friese, Köster, Hassler, Martens, Trujillo-Barreto, and Gruber, 2013; Staudigl and Hanslmayr, 2013; Osipova, Takashima, Oostenveld, Fernández, Maris, and Jensen, 2006; Sederberg, Kahana, Howard, Donner, and Madsen, 2003) and retrieval (Gruber, Tsivilis, Giabbiconi, and Müller, 2008; Osipova, Takashima, Oostenveld, Fernández, Maris, and Jensen, 2006). Functionally, there is good evidence that theta oscillations index the communication between the mediotemporal and neocortical networks, for example, during spatial learning (Jones and Wilson, 2005) and associative memory processes (Backus, Schoffelen, Szebényi, Hanslmayr, and Doeller, 2016; Brincat and Miller, 2015; Kaplan, Bush, Bonnefond, Bandettini, Barnes, Doeller, and Burgess, 2014; Staudigl and Hanslmayr, 2013). A suppression in alpha activity during episodic encoding (Friese et al., 2013; Hanslmayr, Spitzer, and Bäuml, 2009) and retrieval (Klimesch, 1997) is assumed to facilitate cortical processing. More specifically, alpha oscillations index the gating of incoming information by inhibiting task-irrelevant regions and routing it to task-relevant regions (Jensen and Mazaheri, 2010). Furthermore, modulations in the alpha frequency mark states of wakefulness and alertness (Klimesch, 1999). Finally, gamma oscillations (30–100 Hz) in the visual cortex reflect object representation processes (Hassler, Barreto, and Gruber, 2011; Fries, 2009; Bertrand and Tallon-Baudry, 2000), accompanying successful episodic encoding (Friese et al., 2013; Osipova et al., 2006) and retrieval (Gruber et al., 2008; Osipova et al., 2006).

In the present encephalographic (EEG) study, we further investigate sleep-dependent changes in the cortical engram. Neuronal oscillations were analyzed during the retrieval of associative memories pre and post 3 hr of nocturnal sleep, compared with a wake control group (see Figure 4.9). The main focus of the analysis was on the sleep-dependent changes in associative memory circuits during memory retrieval. We thus investigated the changes in neuronal oscillatory activity in the theta, alpha, and gamma frequencies from a pre to a post sleep retrieval phase, compared with a wake group. To see if sleep-dependent changes would be specific to the retrieval of associative memories, sleep-dependent changes for the processing of novel stimuli, forgotten stimuli, and items that were recognized without the associated color were included in

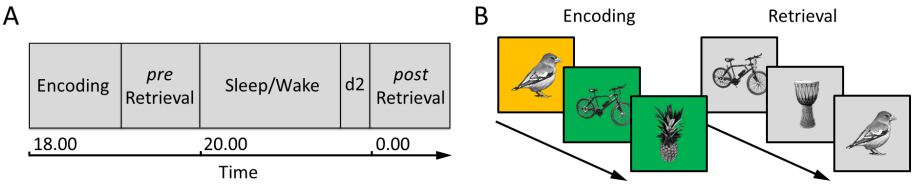


FIGURE 4.9: **The experimental procedure and the stimuli used for the associative memory task.** (A) The experimental schedule of the sessions of the pre–post control group design. The sessions started with the encoding task around 6:00 p.m., after the setup of the EEG apparatus. Participants then performed two retrieval phases, pre and post a sleep (experimental group) or a wake (control group) interval. Participants of the sleep group were equipped with a polysomnography apparatus and went to bed for 3 hr between 8:00 p.m. and 0:00 a.m. The d2 test of attention was applied in both groups, before the second encoding phase, to test for differences in attention and concentration endurance. (B) Associative pictorial memory task. During encoding, participants had to form associations between items and the background color (yellow or green). During retrieval, participants were asked to recognize objects and to recall the associated color among distractor items. Stimulus presentation was identical for encoding and retrieval: Each trial started with a blank screen (1 sec), followed by a fixation point (variable duration of 0.5–1.0 sec) and the presentation of a target stimulus (2 sec). The trial terminated with a question mark that remained until a response was given.

the analysis. Here, a specific focus was on correctly identified novel items, as an indicator for sleep-dependent changes in nonmnemonic processing of visual stimuli.

Methods

Participants

Twenty-six university students voluntarily participated in the experiment, 13 in a sleep group (9 female, $M_{age} = 19.9$ years, $SD_{age} = 2.4$ years) and 13 in a wake control group (9 female, $M_{age} = 20.1$ years, $SD_{age} = 1.6$ years). None of the participants reported any neurological or psychiatric record

or sleep disorders in an anamnesis session. The experimental procedure was conducted in accordance with the World Medical Association's Declaration of Helsinki (59th WMA general assembly, Seoul, 2008), and informed written consent was obtained from each participant. Three additional participants were not included in the analysis because of technical problems (n = 1), insufficient trials for the EEG analysis (fewer than 10 artifact-free trials for item–color hits or correctly identified novel items, n = 1, or a lack of missed items, n = 1), or incomplete data assessments (headache, n = 1).

Participants got out of bed at seven in the morning at the day of the experiment and refrained from caffeine or alcohol consumption on this day (after having one coffee or tea for breakfast). On average, participants reported 7.6 hr of sleep (SD = .9) in the night before the experiment.

Stimuli and Procedure

The stimulus set consisted of 400 gray-scaled pictures of objects (e.g., plants, animals, clothes, tools), taken from a standard picture library (Hemera Photo Objects), presented at a visual angle of 6.2 × 6.2°.

The experimental procedure is depicted in Figure 4.9A. The sessions started around 6.00 PM, after setting up the EEG apparatus. During encoding, participants saw 300 pictures in front of a yellow or a green square (see Figure 4.9B) and were instructed to form object-color associations and where told the example of "a flower on green grass" for a flower that is presented on a green background. To assess the sleep-dependent changes in neuronal activity during retrieval, we employed a pre-post control group design: During encoding participants learned 300 object-color associations. The first retrieval session (pre), followed after a 5 minute break. 150 pictures from encoding were intermixed with 50 new stimuli and presented on a gray background. For each picture, participants had to indicate, whether they retrieved the picture with the associated color (old, green or old, yellow), retrieved the picture without a color (old, no color) or did not see this picture before (new). Here, the response option "old, no color" was included as a response option to avoid extensive color guesses and thereby increase the signal to noise ratio for correctly retrieved item-color associations. A second, identical retrieval session (post) followed around 0.00 AM, after a 180 min. sleep interval (experimental group) or an equally long

awake interval (control group). In the post sleep retrieval session, the second half of the initially encoded stimuli was presented (150 stimuli) along with 50 additional distracters.

Stimulus presentation was identical for encoding and retrieval: each trial started with a blank screen (1 s), followed by a fixation point (variable duration of 0.5-1.0 s), and the presentation of a target stimulus (2 s). The trial terminated with a question mark. Response keys were pressed with different fingers of the right hand. The procedure was demonstrated in 10 training trials prior to each phase of the actual experiment. The stimuli were counterbalanced between sessions (encoding and retrieval, as well as pre- and post-retrieval) equally in both groups (sleep and wake). Object color allocations were randomized for each participant.

Between pre- and post-retrieval sessions, participants were freed from the EEG equipment and had a meal. Participants in the sleep group changed their cloths, were equipped with a polysomnography apparatus (PSG, see Apparatus), and went to bed for three hours between 8.00 PM and 0.00 AM. The PSG was monitored online and participants were only awakened from the sleep stages 1 and 2, to avoid inertia. Participants in the control group saw animated movies, to avoid interference with the real objects used as stimulus material (cf., Alger, Lau, and Fishbein, 2010).

In the sleep group, participants sleep was assessed with an Embla Titanium polysomnography (PSG) system (TNI medical, Würzburg, Germany). The recording montage included six EEG electrodes (F3, F4, C3, C4, P3, P4) and references (M1, M2), such as an electrooculogram (EOG) and an electromyogram at submental muscles. Sleep recordings were classified according to the criteria of the American Academy of Sleep Medicine (AASM; Iber, 2007). The inter-rater reliability between two coders was assessed for 28 random frames of 30 s (Cohens κ = .75).

The d2 test of attention (Brickenkamp, 2002) was applied before the second encoding phase, to test for differences in attention and concentration endurance after the sleep and the wake interval. In this test participants are required to cross out target letters among distractors (every letter d with 2 lines as targets within a series of the letters d, p, or q with 1, 2, 3 or 4 lines) in a limited amount of time. Scores of the test were calculated and compared between groups: total (number of all processed letters minus number of all errors), accuracy (number

of errors), and concentration endurance (number of correct responses minus confusion errors, i.e., mistakenly marked letters).

Sleep-dependent memory consolidation was tested for participants overall memory performance and their associative memory performance, by the means of corrected recognition scores (Pr, i.e., hits minus false alarms; Snodgrass and Corwin, 1988). For the overall memory performance score, all remembered items were counted as hits and misclassified new items were counted as false alarms, irrespective of the color judgement. For participants associative memory performance the proportion of remembered items with correct color judgments (counted as color hits) were corrected by the proportion of remembered items with wrong color judgements (counted as color false alarms, i.e., an estimate for color guesses given that the object was correctly remembered). Both Pr scores were subjected to a mixed-model ANOVA with the factors group (sleep, wake) and retrieval session (pre, post). For all significant main effects and interactions, Greenhouse-Geisser corrected p-values are reported along with the effect sizes, quantified by partial eta squared η_p^2.

EEG Recording and Analyses

The EEG was recorded from 128 active electrodes using a BioSemi Active-Two amplifier system (BioSemi, Amsterdam, Netherlands) at a sampling rate of 512 Hz in a shielded room. A horizontal and vertical EOG was applied to monitor eye movements and blinks. Two additional electrodes (CMS: Common Mode Sense and DRL: Driven Right Leg; cf. www.biosemi.com/faq/cms&drl.htm) served as reference and ground.

Prior to the analysis, continuous EEG data was high-pass filtered at 0.5 Hz and eye-blinks and muscle artifacts were detected using an independent component procedure and removed after visual inspection (Chaumon, Bishop, and Busch, 2015). EEG data was then segmented into epochs from -1000 ms to 3000 ms with regard to the stimulus onset. Further artifacts and noisy trials were removed by the means of statistical correction of artifacts in dense array studies (SCADS; Junghöfer et al., 2000), used in several former studies (e.g. Gruber and Müller, 2006; Gruber et al., 2008). Furthermore, we applied a correction of saccade-related transient potentials (COSTRAP; Hassler, Barreto, and Gruber, 2011), used in several previous publications (Friese et al., 2013;

Hassler, Barreto, and Gruber, 2011; Hassler et al., 2013; Köster et al., 2014) to remove miniature eye-movement artifacts (Yuval-Greenberg et al., 2008). Due to artifact correction procedures approximately 10% of the original trials were removed. Throughout further analysis an average reference was used. To obtain the spectral power over time, the trial data was convoluted using Morlet's wavelets with seven cycles (Bertrand and Pantev, 1994) at a resolution of 1 Hz. To account for the variability in frequency bands across individuals (Klimesch, 1997), we identified the peak theta, alpha and gamma frequencies, individually for each subject, based on the mean spectral activity across the most relevant conditions (item-color associations and correct rejections; cf. Friese et al., 2013). This resulted in (Mean $\pm$ SD): 4.22 Hz $\pm$ 1.53 Hz theta frequencies, 10.92 Hz ± 2.07 Hz alpha frequencies and 60.37 Hz $\pm$ 8.43 Hz gamma frequencies. Event-related spectral changes for each condition and frequency-band were then calculated as the relative signal change of the post stimulus spectral activity, relative to a 500 ms to 100 ms pre-stimulus baseline, in percent. Throughout all further analyses, the relative signal change at individual peak frequencies were used. For all topographical analyses, the relative signal change values at individual peak frequencies were averaged over the full time periode of stimulus presentation (0 ms to 2000 ms), for all comparisons. The main focus of the present study were sleep-dependent changes in associative memory processes. Furthermore, to test whether or not sleep-dependent changes in EEG activity would be specific to the retrieval of associative memories, we also included the trials of correctly identified novel items, hits without color judgement and missed items.

Grand mean oscillatory activity. In a first step, we analyzed the grand mean spectral changes relative to the baseline in the first retrieval session, namely prior to the group-specific experimental manipulation (sleep or wake). Values at each electrode were averaged over all trials of the conditions entered included in the analyses. The signal changes for individual theta, alpha and gamma frequencies were tested against baseline using a cluster mass permutation t-test (two-sided, dependent samples; Maris and Oostenveld, 2007). To further characterize the event-related signal changes over the stimulus presentation time, we calculated the grand mean activity at each time point, at frequency specific electrode clusters, i.e., at the most significant clusters for theta, alpha, and gamma.

Sleep-dependent changes in neuronal oscillatory activity during memory retrieval. The sleep-dependent changes in retrieval activity were analyzed, following the rationale of the pre-post control group design: The pre-post differences in the three frequency bands (theta, alpha, gamma) were compared between the wake and the control group, to obtain the sleep-specific changes in neuronal oscillatory activity. Specifically, we applied a two level statistical analysis (cf. Obleser et al., 2012). At the first or subject level, we obtained t-values for the trial data of individual pre and post differences using an unpaired t-test (testing for significant differences, two-sided, independent samples). This was done separately for the four conditions included in the analysis (item-color hits, novel items, hits without color and misses). Attaining first-level statistics rather than using only average power changes compared to baseline has the advantage that the pre-post differences that are subsequently entered into group statistics are effectively standardized for across-trial variances and differences in the number of trials between subjects and conditions. At the second or group level, the individual t-maps of all four conditions were used as the input of a combined cluster mass permutation t-test (testing for significant differences, two-sided, independent samples; adapted from Maris and Oostenveld, 2007; implemented in fieldtrip, Oostenveld et al., 2010). This was to check for significant differences in changes from pre to post (first level) between the sleep and the wake group (second level). Specifically, the cluster statistic was calculated as the sum over the t-values of neighboring electrodes with a cluster inclusion criterion of $p < .05$, separately for each conditions included in the analysis. The significance of the cluster statistics was then calculated from a combined permutation distribution obtained from 1000 Monte Carlo iterations with randomly assigned sleep and wake groups, including the clusters of all four conditions. The results are cluster wise p-values which are not affected by inflated false-positive rates otherwise arising from multiple comparisons. The rationale for including all four conditions into a combined permutation, was to consider the whole variation for pre-post differences when determining clusters that exceed chance level.

TABLE 4.1: **Attention and Concentration Endurance.**

d2 score	Sleep	Wake	t	p
Total	459.1 (75.6)	496.8 (58.3)	-1.43	$> .15$
Accuracy	15.1 (7.6)	23.3(19.3)	-1.44	$> .15$
Concentration Endurance	184.5 (28.3)	187.6 (57.5)	$-.18$	$> .25$

Means $\pm$ SD are displayed. P-values denote differences between the sleep and wake group (results of pairwise comparisons, two-sided). Total: number of processed letters minus number of errors, Accuracy: number of errors, Concentration endurance: number of correct responses minus confusion errors.

Results

Behavioral Results

Subjects in the sleep group showed normal sleep profiles for an early night sleep (Mean ± SD): The total sleep time was 137.9 ± 20.0 min with pronounced episodes of slow-wave sleep (stage 1 sleep [S1]: 14.3 ± 12.2 min; stage 2 sleep [S2]: 51.4 ± 19.9 min; slow-wave sleep [S3]: 66.5 ± 31.9 min; rapid eye-movement sleep [REM]: 9.2 ± 11.3 min; wake after sleep onset: 25.8 ± 39.7 min).

The results of the d2 attention and concentration endurance scores are reported in Table 4.1. There were no differences in the total score, the accuracy or concentration endurance score between the sleep and the wake group.

Over both groups (sleep and wake) and conditions (pre and post), subjects remembered (Mean ± SD) 77.6 ± 8.7 % of the target pictures (hits) while judging 21.5 ± 8.7 % as new (misses). Of the distracters 78.5 ± 15.2 % were correctly recognized as new (correct rejections) and, accordingly, 22.4 ± 15.2 % were classified old (false alarms). Regarding the retrieval of object-color associations, 46.9 ± 12.8 % of all target pictures were recollected with the correct background color (hit, color), 16.5 ± 7.3 % were retrieved with false color judgements (false alarm, color) and 13.7 ± 5.4 % were retrieved without color judgement.

TABLE 4.2: **Overall and Associative Memory Performance.**

	Pre	Post	t	p
Memory Performance (P_r)				
Sleep	.59 (.23)	.53 (.16)	1.31	$> .20$
Wake	.66 (.12)	.47 (.14)	5.44	$< .001$
Associative Memory Performance (P_r)				
Sleep	.30 (.24)	.21 (.13)	1.77	$> .10$
Wake	.36 (.10)	.21 (.08)	4.99	$< .001$

Means $\pm$ SD are displayed. P-values denote significant changes between retrieval phases (results of pairwise, two-sided comparisons).

For the memory performance of the sleep and the wake group in the pre and the post phase, see Table 2. Overall memory performance showed a decay over time, F(1, 24) = 19.88, p < .001, η_p^2 = .453, and a sleep-dependent consolidation effect. This is indicated by a significant phase × group interaction, F(1, 24) = 5.88, p = .023, η_p^2 = .197, as well as a significant decay in memory performance in the wake, but not in the sleep group (see Table 2). Associative memory performance also decreased from pre to post retrieval, F(1, 24) = 16.05, p < .001, η_p^2 = .401. While the phase × group interaction was not significant, a consolidation effect is indicated by a reduction in the wake, but not in the sleep group (see Table 2). Although baseline differences between the sleep and the wake group contribute to these effects, baseline performances for overall and the associative memory did not differ significantly between groups, t = -1.03, p = .32, and, t = -.77, p = .45.

EEG Results

Grand mean oscillatory activity during retrieval. The signal changes in oscillatory power in the first retrieval phase revealed significant increases in theta power at frontal (cluster test against baseline: p < .001) and posterior (p = .016) electrodes. At frontal electrodes, theta activity reached a peak around 600 ms after stimulus onset. Overall alpha power was suppressed across all electrodes (p < .001). Relative to the theta increase, the onset and the peak of the alpha suppression were slightly delayed (100 ms). The gamma power showed an increase in the occipital region (p = .050) with a peak around 300 ms

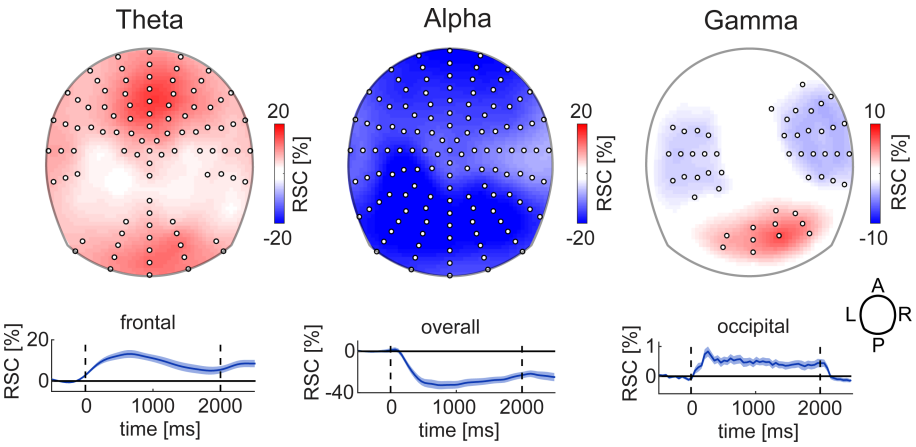


FIGURE 4.10: **Grand mean spectral power in the theta, alpha and gamma frequencies during retrieval.** Topographical maps of the relative signal changes (RSC) in grand mean activity for item-color hits, correctly identified novel items, hits without color judgements and misses. Topographies show the averaged power for the whole duration of the stimulus presentation (0 ms to 2000 ms) at individual theta, alpha and gamma frequencies. Only significant differences ($p < .05$) are displayed. White dots indicate the electrodes of statistically significant clusters ($p < .05$), compared to baseline. Lineplots show the temporal evolution of the event-related signal changes per time point, for the frontal (theta), overall (alpha) or occipital (gamma) cluster. Shaded areas are 95% confidence intervals.

and a decrease to baseline level after stimulus offset. We furthermore found unexpected event-related decreases in temporal gamma activity (both $p < .021$). Grand mean event-related changes in the pre retrieval phase are displayed in Figure 4.10. Notably, a comparison of the pre retrieval oscillatory activity between the sleep and the wake group, confirmed that there were no differences in retrieval activity between both groups. This is, there were no significant clusters for any frequency band in any of the conditions included in the analyses (all $p > .116$).

Sleep-dependent changes in neuronal oscillatory activity during memory retrieval. To test the sleep-dependent changes in the retrieval activity, we

compared the time-dependent changes (post-pre) between the sleep and the wake group (Figure 3B) for all conditions included in the analysis. For correctly retrieved item-color associations, we found reduced event-related theta activity after the sleep interval and an increase in event-related theta activity after the wake interval (Figure 4.11A, left column). Group differences were found in a significant cluster in the posterior region ($p = .024$) and in a cluster over the left temporal-posterior region ($p = .015$), see Figure 4.11B and Figure 4.11C, left column. The event-related alpha suppression was reduced in the wake group while the sleep group maintained the same level of alpha suppression. This group effect in the alpha frequency is significant in frontal ($p = .011$) and parietal ($p = .041$) clusters. Sleep-dependent changes in the gamma frequency were found at posterior electrodes, namely the electrodes with the highest gamma power in the grand mean (see Figure 4.10). The sleep group showed a decrease in gamma power while the wake group showed increased gamma (Figures 4.11 A-C, right panel). The parietal cluster was not significant in the cluster mass permutation test ($p = .111$), possibly because of the very focalized spatial distribution of parietal gamma power, that does not survive the cluster permutation test. However, in a data driven approach that evaluates the occipital electrode cluster or the peak gamma electrode of the grand mean, resulted in significant differences between the sleep and the wake group (cluster: $t = 2.29$, $p = .041$; peak electrode: $t = 2.51$, $p = .027$). Note that sleep-dependent changes in gamma were not found in the further conditions included in the analyses (cluster: all $p > .201$; electrode: all $p > .267$).

We further focused on the processing of novel stimuli to see whether sleep would likewise modulate the processing of novel stimuli. Sleep-dependent changes in the processing of novel stimuli did not reveal event-related changes in theta and gamma activity, but in alpha power (Figure 4.12A). Namely, there was a significant reduction in event-related alpha suppression from pre to post in the wake group but not in the sleep group (frontal cluster: $p = .004$; parietal cluster: $p = .034$; Figure 4.12B and Figure 4.12C). For the further conditions, hits without color judgment and misses, there were no significant differences between the sleep and the wake group, but a marginally significant reduction in frontal alpha suppression ($p = .069$) after the wake compared to the sleep interval (Figure 4.13).

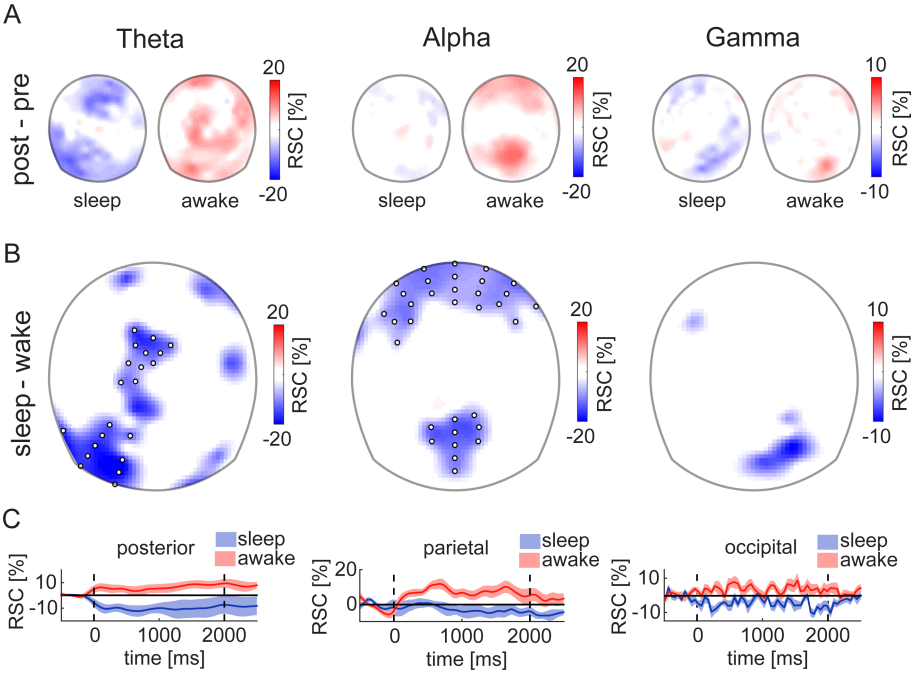


FIGURE 4.11: **Spectral power for the retrieval of item-color associations.** (A) Topographical maps of the time-dependent changes in retrieval activity from pre to post (post - pre) for the sleep and the awake group. Topographies depict relative signal changes (RSCs) for the whole time of the stimulus presentation (0 ms to 2000 ms). (B) Topographical maps show the sleep-dependent changes for retrieval activity for item-color associations, following the rationale of the pre-post control group design: The pre-post differences, as depicted in panel A, were compared between the wake and the control group (sleep - wake). Only significant differences ($p < .05$) are displayed. White dots indicate the electrodes of statistically significant clusters ($p < .05$), identified by a cluster mass permutation test. Note that sleep-dependent changes in gamma power were significant at the electrodes of the grand mean parietal gamma cluster ($p = .041$). (C) Time series show the averaged post-pre differences of the sleep group (blue) and the wake group (red), averaged over the electrodes in the clusters of panel B. Shaded areas are 95% confidence intervals.

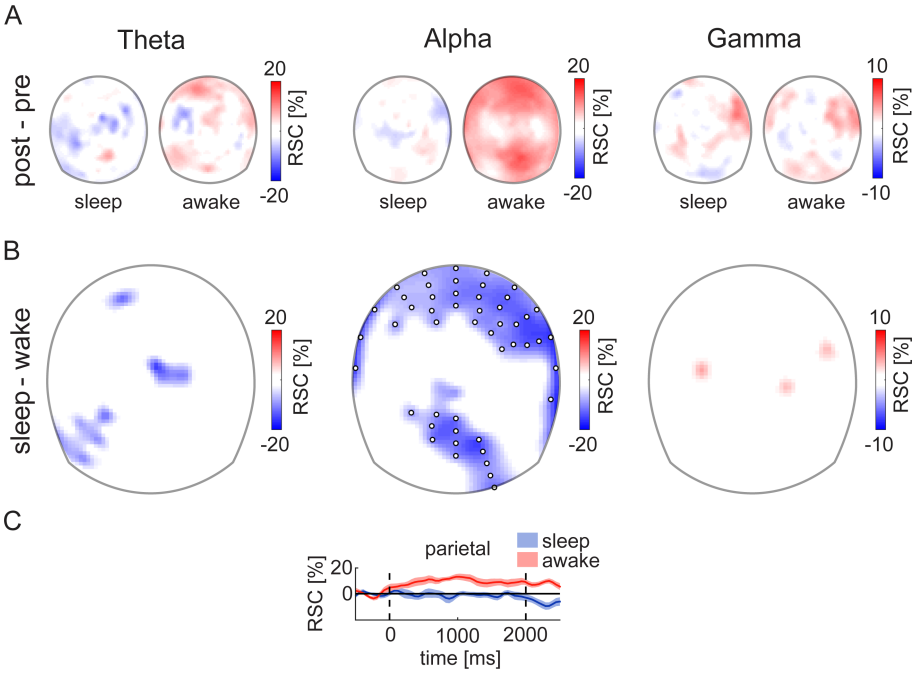


FIGURE 4.12: **Spectral power for items retrieved without color.** (A) In parallel to the analysis of the retrieval of item-color associations (see Figure 4.9) topographical maps depict time-dependent changes (post - pre), for correctly identified novel items in the sleep and the awake group. Topographies depict the relative signal changes (RSCs) between 0 to 2000 ms. (B) Sleep-dependent changes: The pre-post differences, as depicted in panel A, were compared between the wake and the control group (sleep - wake). Only significant differences ($p < .05$) are displayed. White dots indicate the electrodes of statistically significant clusters ($p < .05$). (C) Time series show the averaged post-pre differences of the sleep group (blue) and the wake group (red), averaged over the electrodes in the parietal cluster of panel B. Shaded areas are 95% confidence intervals.

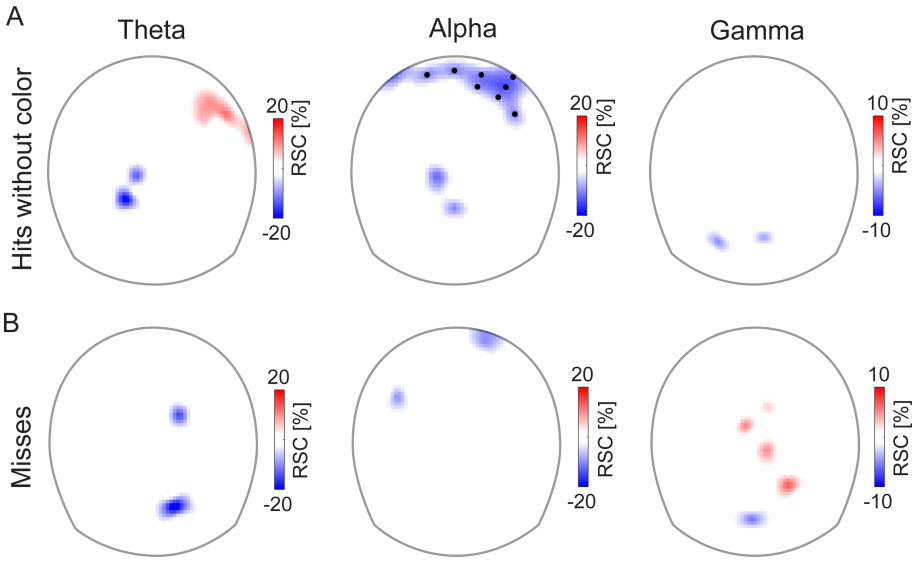


FIGURE 4.13: **Spectral power for items retrieved without color and misses.** Topographical maps depict sleep-dependent changes for (A) items retrieved without color and (B) items that where missed during retrieval. The pre-post differences (post-pre), were compared between the wake and the control group (sleep - wake), compare panel B of Figures 4.11 and 4.12. Topographies depict the relative signal changes (RSCs) between 0 to 2000 ms. Only significant differences ($p < .05$) are displayed. Black dots indicate the electrodes of the only cluster that was marginally significant in these two conditions ($p = .069$).

Discussion

We analyzed neuronal oscillations in the human EEG during the retrieval of associative memories before and after a sleep, compared to a wake control group, to investigate sleep-dependent changes in the cortical engram. Sleep-dependent memory consolidation was indicated by the behavioral data and specifically modulated associative retrieval processes reflected in the theta and the gamma frequency range: Event-related theta power was reduced after a sleep, in contrast to a wake interval, peaking at centro-parietal electrodes. Likewise, occipital gamma activity was reduced after a sleep but increased after a wake period. In contrast, we did not find sleep-dependent changes in the theta and the gamma frequency for the processing of novel stimuli, hits without color judgements or misses. Event-related alpha suppression for the retrieval of associative memories was reduced from pre to post in the wake group for the processing of novel stimuli and was also marginally reduced for items that were recognized without color, but not for missed items. Thus, the retrieval of associative memories after sleep was specifically associated with a reduction in theta and gamma activity, while a reduced alpha suppression after sleep seems to be a more general effect of sleep on the processing of visually presented stimuli.

The theta rhythm is a prominent marker of associative memory processes in the medio-temporal lobe (Fell and Axmacher, 2011; Lisman and Jensen, 2013). Furthermore, theta oscillations in neocortical neuronal networks were found to synchronize with theta oscillations in the medio-temporal lobe during associative retrieval processes in the monkey (Brincat and Miller, 2015) and human brain (Backus et al., 2016; Kaplan et al., 2014). The theta rhythm is assumed to index the interplay between medio-temporal and neocortical neuronal networks (Brincat and Miller, 2015; Kaplan et al., 2014; Köster et al., 2014; Nyhus and Curran, 2010). Here, we found reduced theta activity during memory retrieval after a sleep interval (Tononi and Cirelli, 2006). Reduced theta activity may possibly index a reduced and possibly more efficient interplay between the medio-temporal and the cortical memory system, which would be in line with the active system account of memory consolidation (Born and Wilhelm, 2012). In particular, a reduced or more effective interplay between the medio-temporal and the cortical memory system may result from a partial transfer of information between both structures during sleep. Recent findings from fMRI

research sugessts that the MTL and neocortical engram is reactivated during sleep (Deuker et al., 2013) and facilitates the communication between MTL and neocortical networks in the human brain (Gais et al., 2007). Occipital gamma activity may reflect neuronal synchronization processes accompanying visual perception of objects (e.g., Osipova et al., 2006; for a review, see Tallon-Baudry and Bertrand, 1999), which have been shown to be sensitive to memory retrieval processes (Osipova et al., 2006; Köster et al., 2014). The reduction in occipital gamma activity has formerly been described for repeatedly presented objects (Busch et al., 2006; Hassler, Barreto, and Gruber, 2011; Hassler et al., 2013) and is interpreted as a reduction in neocortical activity, due to sharpened or globally reduced synaptic activity for the processing of stimulus material (Wiggs and Martin, 1998). Here we found a sleep-dependent reduction in occipital gamma during the retrieval of associative memories, as indicated by significant differences in the parietal cluster from the grand mean, but not in the cluster test. Both findings, i.e., reduced theta and gamma activation for post sleep retrieval thus conform with a synaptic homeostasis view on sleep-dependent memory consolidation, namely that sleep promotes a global downscaling of synaptic strength, leading to a more effective processing in recently activated memory traces (Tononi and Cirelli, 2006). Noteworthy, a recent study by Kuhn et al. (2016), also found decreases in theta and gamma activity after sleep, which is also interpreted as an indicator of more general downscaling processes in the human cerebral cortex after sleep.

Critically, the reduction in theta and gamma activity was selectively observed for associative memories, but not for items retrieved without color. This substantiates the idea that theta and gamma neuronal oscillations index specific neuronal mechanisms that underpin the initial formation of lasting associative memories within the MTL and neocortical memory system (Lisman and Jensen, 2013), which is also assumed to play a key role in the consolidation of associative memories during sleep (Diekelmann and Born, 2010; Rasch and Born, 2013).

Modulations in alpha activity have been shown to reflect diurnal variation in alertness and wakefulness (Klimesch, 1999). Event-related alpha suppression is a marker of attentional processes (Berger, 1929) and is thought to

gate task relevant processes in neocortical networks (e.g., Jensen and Mazaheri, 2010; Klimesch, 1999; Klimesch, Sauseng, and Hanslmayr, 2007). Research suggests that event-related alpha suppression during mnemonic processing is rather associated with semantic processes accompanying visual perception processes (Hanslmayr, Spitzer, and Bäuml, 2009). In the present study, modulations in alpha activity were found in the wake group (where alpha suppression was reduced after sleep) and were not specific to memory retrieval but also found for novel stimuli and hits without color judgments. Thus, reduced alpha suppression may reflect decreases in attentional resources and the semantic processing of the stimuli due to the three-hour period of wakefulness. These findings suggest that changes in the alpha band do not index processes which are specific to associative memory retrieval, but may support the idea that sleep also promotes mnemonic functioning by a recovery of attentional resources. These may also contribute to a lower decay of memory performance found in the sleep group. This interpretation leaves open the question, why there were no sleep dependent changes in alpha oscillations for missed items. Furthermore, the d2 test of attention and concentration endurance did not reveal any differences in attentional resources after the sleep or the wake interval, indicating that there was no severe effect of wakefulness on subjects' task performance. To summarize, our findings substantiate the idea that sleep-dependent memory consolidation is a multifaceted phenomenon, relying on multiple neuronal mechanisms and processes, working in concert.

To conclude, to our best knowledge, this is the first study that analyzed neuronal oscillatory activity during memory retrieval to investigate sleep-dependent changes in the cortical memory trace. Most importantly, we found a sleep-dependent reduction in theta activity, indicating a reduced or sharpened interplay between MTL and neocortical networks and a reduction in occipital gamma, indicating reduced synaptic activity in visual cortex. These findings highlight the notion that sleep leads to a marked transformation of cortical memory representations, presumably driven by active system consolidation and synaptic homeostasis processes, working in concert. However, the results of the present study should be substantiated in future studies and cross-validated with other methodologies, such as MEG or intracranial recordings. Methodologically, the present study further illustrates that pre-post sleep comparisons are a promising approach to further elucidate the transformations that memory representations undergo during periods of sleep.

4.3 Chapter Summary

The two manuscripts presented in this chapter demonstrate the importance of dynamic organization of cortical multi-site communication for behavioral and cognitive functions. These multi-site interactions in the brain were investigated with a multisensory integration task in the first study and with an associative memory task in the second study. Here, I will place these findings into the context of this thesis.

The first manuscript showed that the behavior of subjects in the sensory integration task with the sensory augmentation device is best described by a subjective Bayesian alternation model. More specifically, some subjects showed an overall preference for one of the two modalities, while other subjects where able to combine information from both modalities. The Bayesian alternation model is able to account for both types of behavior. Therefore it can explain the empirical data better in comparison to a sole Bayesian integration model which presumes the optimal integration of both modalities.

These experimental results can be further interpreted in relation to the topic of structural, functional and dynamic brain interactions. The Bayesian alternation model can explain that subjects switch between using the vestibular sense and using the tactile augmentation device. This indicates that the communication between brain areas is dynamically coordinated such that the information about angular rotation can be obtained either from the tactile, or the vestibular, or a combination of both senses. This dynamic coordination can be assumed to be resulting from cortical synchronization of the involved brain areas, as it was hypothesized and illustrated in Figure 1.10A in the general introduction of this thesis. Furthermore, one can speculate that the unsupervised learning, which was presented in chapter 3, could account for the learning of a mapping between the newly augmented tactile sense and the vestibular sense. Applying an autoencoder learning rule could lead to the development of optimal integration behavior, mediated in sensory association areas as illustrated in Figure 1.10B.

In the second manuscript presented in this chapter, we investigated oscillatory brain activity in the context of the consolidation and retrieval of associative memory items before and after sleep. In particular, we recorded and analyzed sleep dependent changes in oscillatory activity during the retrieval

period. Changes in brain oscillations between the pre-sleep and post-sleep period indicate that the cortical representations of associative memory items were transferred from medial temporal to neocortical networks. Specifically, the reduced theta activity during memory retrieval after sleep could be interpreted as a reduced involvement of the hippocampus after the manifestation of memory representations in statically wired connectivity within the neocortex.

Again, these experimental results can be framed in the context of normative learning of memory representations as presented in chapter 3. For example, the transfer of associative memory items from the hippocampus into neocortical networks during sleep could be caused by the repetitive activation in the hippocampus triggering gradual changes in the neocortical synaptic weights. Furthermore, the manuscript highlights the overall importance of oscillatory activity for mediating the communication channels between the hippocampus and neocortical networks. Overall, the results show that changes in the structural connectivity due to learning lead to changes in the dynamic oscillations in the brain during retrieval of these learned memory items. This again highlights the important interplay between learning of structural connectivity on slow timescales and dynamic organization of functional brain activation on fast time scales.

Chapter 5

General Discussion

In this thesis I investigated several aspects of the interplay between structural connectivity, functional connectivity, and fast dynamic oscillatory interactions. The first manuscript showed how a computational model based on structural brain connectivity can predict empirical functional brain connectivity. Afterwards, the second manuscript focused on the computational principles, which are implemented by these neurophysiological mechanisms, by investigating models of normative learning from natural sensory inputs and the contextual coding provided by the coupling of neural phase oscillators. The last two manuscripts broadened this view by analyzing the cognitive abilities that are facilitated by these neurophysiological mechanisms. In particular, one manuscript investigated sensory integration or alternation and the last manuscript analyzed oscillations in relation to memory functions.

Overall, all presented studies demonstrated how dynamic coordination bridging multiple timescales facilitates the flexibility of information processing in the human brain. Specifically, the results of the connectome modeling study show that a model that includes dynamic time delayed interactions is a better predictor of empirically measured functional brain connectivity in comparison to a simpler model of only stationary dynamics (Finger et al., 2016). Moreover, the temporal coding study highlights that the combination of representation learning on slow timescales, and network activations by sensory stimulation on medium timescales, and dynamic binding by synchrony on fast timescales can be combined in a single computational model. This model that bridges all timescales is able to bind neural representations of natural visual images and segment these images in a network of coupled phase oscillators (Finger and

König, 2013). Next, the sensory integration study highlighted the important functional role of coordinating large-scale brain communication to integrate information from multiple sensory areas (Goeke et al., 2016). The results of this experimental study show that participants alternated between the native vestibular and the augmented rotation information from the tactile modality indicating that the communication between brain areas is dynamically reorganized. Finally, the memory consolidation study demonstrates that dynamic coordination in the brain is specifically important for the consolidation of associative memories (Köster et al., 2016). The results showed sleep-dependent memory consolidation in the behavioral data and that sleep modulated the theta and the gamma frequency range specifically for associative retrieval processes.

More detailed summaries of the corresponding findings were already provided in the general introduction and in the summary sections of the respective chapters. Also the relation to previous work was already extensively explained in the general introduction (chapter 1) and in the individual papers. Thus, to avoid unnecessary repetition, the remainder of this general discussion chapter will focus mostly on the overarching consequences of the presented work, the prospective outlook for future work, and new ideas that were developed during my work. These new ideas might be helpful to guide further work in this interesting line of research.

5.1 STDP-Learning, Oscillations, and Autoencoders

In this thesis, I presented microscale models of learning in autoencoder networks and macroscale models of the cortical connectome. To unify these two approaches, each autoencoder layer could be thought of as implementing one specific cortical module, or as one specific node in the connectome. Following this perspective the forward and backward connections of the autoencoder can be interpreted as the edges in the connectome measured as white matter connectivity. This poses the question if one could further unify these concepts of unsupervised learning with autoencoders and the concepts of oscillatory couplings within and between cortical modules. In this section, I will discuss how these concepts could be related by considering how precise timings of action potentials might be important for learning and coordination of brain functions.

Spike-timing dependent plasticity (STDP) is a neurophysiologically realistic model describing the change of synaptic weights based on the activation statistics of the pre- and post-synaptic neurons. It seems reasonable to suppose that oscillatory network activations facilitate STDP learning by inducing a temporal reference frame for the timing of the pre- and post-synaptic action potentials (Paulsen and Sejnowski, 2000). Oscillations in biological networks can be produced by inhibitory interactions in response to a period of excitation. In addition, a constant input pattern to such a network of oscillating integrate-and-fire neurons naturally orders the spike timings in a way that is repeated across consecutive oscillation cycles. Rumbell and colleagues used this mechanism in a spiking implementation of self organizing maps (SOM), where STDP rules were shown to facilitate temporal structuring in oscillatory neural activations (Rumbell, Denham, and Wennekers, 2014). One might ask, if a similar mechanism could be used by the brain to compress the high bandwidth sensory inputs into more meaningful representations with a computational principle resembling autoencoders.

Recently, it was shown that a network with integrate-and-fire neurons and STDP learning rules can implement a global autoencoder objective function (Burbank, 2015). This is remarkable, because autoencoders are normally implemented by back-propagation of error gradients, which requires a neurophysiologically unrealistic distant effect. Similar to Rumbell and colleagues, Burbank used an inhibitory pool within each layer so that the action potentials

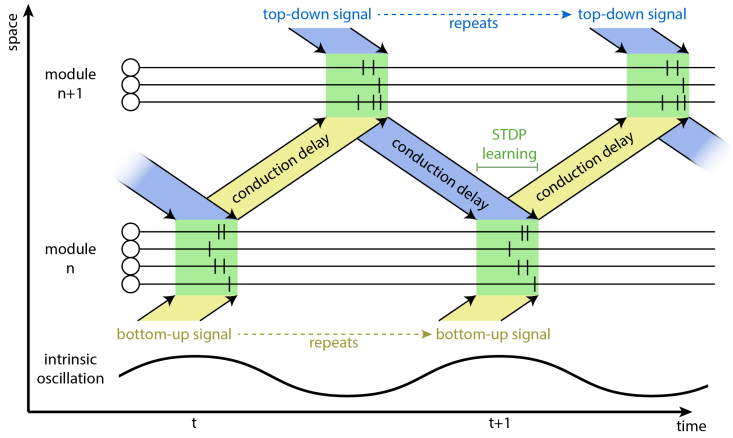


FIGURE 5.1: **Coordination of STDP learning by synchronized oscillations.** If the arrival of top-down inputs is just prior to the spikes in the next oscillation cycle of the lower layer, then the learning will strengthen synaptic weights to decode the hidden layer representation in a way that reconstructs the input signal.

in a local region are phase locked to the overall oscillation of that region. In this regard, one could further speculate that the intrinsic oscillation is an essential element that facilitates STDP learning between cortical modules and thereby allows the optimization of an autoencoder objective function. Figure 5.1 illustrates the idea of this mechanism in a space-time diagram by depicting the bidirectional information transfer between two distant cortical modules. Considering intrinsic oscillatory neural activations, one can assume that input spikes arrive in a temporal pattern that is similar across consecutive oscillation periods. Therefore, the intrinsic oscillations create a local common time reference in each cortical module with spike timings locked to the phase in the local-field potential.

The STDP learning rule can be described in a simplified way, specifically as strengthening the weights of synapses, which contributed to the activation of the post-synaptic neuron. For simplicity, let us assume that the conduction delay between two cortical modules corresponds to roughly half of the intrinsic oscillation cycle and that the local field potentials of the two cortical modules are antiphasic as illustrated in Figure 5.1. Then the bottom-up (yellow) and

top-down (blue) signals would arrive roughly at the right time to contribute to the activation of the corresponding repeating spike patterns. Therefore, both these input signals are integrated during a short time frame (marked in green). The resulting action potentials would evoke STDP learning such that the top-down and bottom-up inputs that contributed to the activation are strengthened.

To further explain how these forward and backward projections between the two modules implement the minimization of the reconstruction error, we further assume that the external inputs are relatively stable over time. More precisely, the bottom-up input to module n (yellow input at the bottom of the figure) and the top-down input to module n+1 (blue input at the top of the figure) are repeating the same patterns across consecutive oscillatory periods. Under this assumption, the top-down connections to module n at time t+1 (blue region in the center of the Figure) would be modified by STDP learning rules to reconstruct the original bottom-up input of module n (yellow) that was received at time t and was forward propagated in the previous cycle. The same argument can be applied for the learning of the forward connections, assuming a stable repeating top-down input from above. To conclude, cortical oscillations might facilitate the learning process of an autoencoder objective function in spiking networks.

At the same time these intrinsic oscillations might provide a mechanism to organize bottom-up and top-down interactions. In particular, one could further ask the question how these cortically embedded autoencoders interact with brain oscillations. Interestingly, the described architecture would also allow dynamic self organization by opening or closing communication channels between specific pairs of modules (Fries, 2005; Womelsdorf et al., 2007; Fries, 2009; Bastos, Vezoli, and Fries, 2015). For example, if the oscillations within the upper and lower modules are desynchronized, then the communication channel between these two modules would be closed so that no information is transmitted and also no unsupervised learning would take place. This opening and closing of the communication channels could also be described as a gating of the information flow in the network as shown in Figure 5.2. This dynamic modulation of the effective connectivity could take place globally between all neurons of one module to all neurons within another module as described with the autoencoder above. But similarly also the effective connectivity between specific neuronal populations of one module and specific neurons in another module might be modulated as described in the following paragraph.

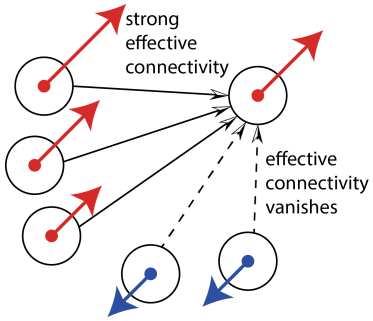


FIGURE 5.2: **Gating of information flow.** The phase of network nodes is depicted by the direction of the arrows. The effective functional connectivity between the nodes depends on their relative phases.

In this thesis, the phase interactions that code binding of neural activations were implemented as horizontal connections that couple Kuramoto oscillators within one of these cortical modules. In a follow-up project we further investigated the ability for temporal coding by phase variables across modules in the framework of complex valued autoencoders. The separation of the functional activation of each neuron into one rate variable and one phase variable can easily be represented as a complex number. This complex number has an absolute value indicating the firing rate and a complex argument (angle) indicating the phase of firing. Furthermore, also the connection weights of an artificial neural network can be represented as a complex number. In this setting, the absolute value indicates the connection strength, while the complex argument can additionally represent the conduction delay, which in this framework corresponds to the preferred phase offset. This leads to the fact that a normal matrix multiplication of the complex valued weights with complex valued input activations results in the correctly delayed output activation, with a correspondingly shifted phase offset. To conclude, these ideas about the relations between autoencoders and the cortical neurophysiology could guide further analysis of the functional role of cortical oscillations and their phase synchronizations.

5.2 Spatial Scales and Levels of Abstraction

The presented studies investigated the interplay between brain structure and brain function from different perspectives. In the first study, structural and functional brain connectivity was mostly viewed on a descriptive level, without a deeper analysis of their computational roles or the behavioral consequences. The focus on a relatively large spatial scale allowed the empirical assessment of structural and functional connectivity from human participants. In contrast, the second study focused on the computational principles of the involved neural processes, which was accomplished using a microscopic view only on the visual subsystem to model unsupervised learning from the activation statistics. Finally, the last two manuscripts highlighted the behavioral consequences and therefore focused on an even broader perspective by analyzing the response behavior of participants, in the form of their estimation of the rotation angle or their answer to a memory retrieval task. This poses the question, how these different spatial scales and levels of abstraction could be further unified into a single theory or a single computational model. In this regard, we started further projects that investigate methods to integrate the macroscale and the microscale model.

The structural connectome that was used in the first manuscript was obtained by a parcellation into 66 regions defined by the freesurfer tool. However, the diffusion tensor data is available in a much higher resolution with 2 mm voxel diameters. Therefore, it makes sense to use a parcellation scheme that results in a connectivity matrix that includes these finer details. In this regard, in an ongoing project we use fiber tractography results on the voxel level resulting in a connectivity matrix between all pairs of 42'000 gray matter voxels. This high resolution raw connectivity matrix can then be clustered with a normalized cut objective function (Shi and Malik, 2000). In the future, this higher resolution connectome could allow a more detailed computational model possibly explaining more variance in the empirically recorded functional connectivity.

Deep hierarchical models that learn from sensory statistics are usually defined as a linear sequence of layers that are homogeneously stacked. This architecture gives rise to larger and more abstract features in higher layers that are based on more simple features in lower layers. In comparison, the human connectome is not organized in a linear sequence, but connects the cortical regions

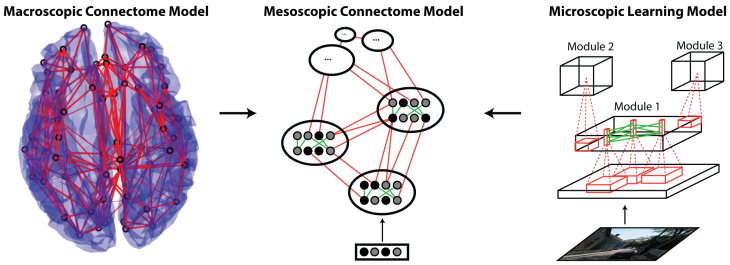


FIGURE 5.3: **A mesoscopic connectome model.** Higher resolution diffusion tensor imaging data could allow specific initialization of a mesoscopic connectome architecture (as described in chapter 2), while normative models of learning could fine tune the connection weights within and between modules (as described in chapter 3).

probably in a "small world" type of network (He, Chen, and Evans, 2007). A higher resolution parcellation of the connectome, as described above, could be used to setup a hierarchical modular network with multiple interacting regions in a highly non-linear network topology that is more similar to the connectivity in the human brain. Figure 5.3 illustrates such a model on a mesoscopic spatial scale. This normative model of unsupervised learning, which is constrained to an architecture informed by diffusion tensor imaging, could be interpreted in the following way: The information that is extracted from the coarse structural connectome could shape the architecture of the network similar to how genotypic information guides the development of the human brain. On the other hand, the activation statistics of the individual network nodes could shape the detailed individual connection weights similar to how the phenotype influences the development of the human brain. Therefore, the structural connectome would represent the genotype while the normative model of learning from natural sensory stimuli would represent the phenotype in such a model.

In the visual system, the cortex is organized in a retinotopic way (Engel, Glover, and Wandell, 1997). The diffusion tensor data also provides new opportunities to analyze these topological properties in further detail. A future project could use the higher resolution fiber tractography results to classify the connectivity between regions into different topological projection types. Specifically, the connection between two cortical modules could be classified

based on the mapping from the internal topological structure of one region to the other. For example, some connections might project in a parallel or mirror-symmetric way and thereby preserve the topological space. In contrast, other connectomic connections might project in an all-to-all fashion from one module to another. These types of connections could be translated to convolutional (topology preserving) and dense (all-to-all) layers in deep learning architectures. However, convolutional connections are based on the assumption that features are translation invariant (the same neural weights are applied as a filter kernel at all positions in an image). This leads us to a third projection type that is usually not used in deep learning architectures, but could be very useful in the future: Connectivity from one layer to another could be constrained to a local neighborhood without the need to tie the weights between all locations. In summary, there are many ways in which diffusion tensor imaging could be further used in the setting of unsupervised learning of artificial neural networks to precisely specify the topological mapping between modules.

5.3 Dynamic Coordination in Deep Learning

I presented a model of functional connectivity on the full brain level that is informed by empirical data and another model of functional connectivity on the microscopic scale, which is learned bottom up. The previous section discussed a possible combination of both models by increasing the resolution of the connectome. One might ask, if the stacking of autoencoder modules could also result in network architectures that are more similar to the human connectome. This bottom-up approach by normative learning from input statistics could make use of phase variables as described in chapter 3. These phase variables could then dynamically coordinate activations not only within modules but also across multiple modules.

In this regard, Figure 5.4 visualizes how such a deep model could utilize the phase variables for dynamic organization in multi-site communication. The proposed architecture uses autoencoders throughout several modules that learn unsupervised from natural sensory statistics. On the bottom, there are two modules depicted that are responsible for two different sensory modalities, such as the primary somatosensory area for proprioception and primary visual area for vision. For simplicity, the feature space of both these modalities is depicted as a 1-dimensional space. The depth axis depicts the learned feature dimension for each autoencoder layer that is stacked on top of the sensory inputs. The Figure shows an exploded-view drawing with each autoencoder layer shifted apart, such that the within layer activations (highlighted nodes and edges) and phase variables (color: red or green) are visible. As seen in the Figure, the feature dimension of the two top layers of the hierarchical model branches out into two streams, that could resemble the ventral and dorsal pathways of the human brain. Overall, this visualization highlights the idea how sensory modalities could be integrated by higher layers, and importantly, how separate higher layer pathways might develop by a branching of the feature dimension of the stacked autoencoders.

This architecture also allows to reason about possible implications and use cases of the phase variables that were presented in chapter 3. As an example, consider a human holding a red book in the right hand, as shown in the figure. If there is input in both modalities that originates from the same physical source, then the representations from both modalities can be integrated in

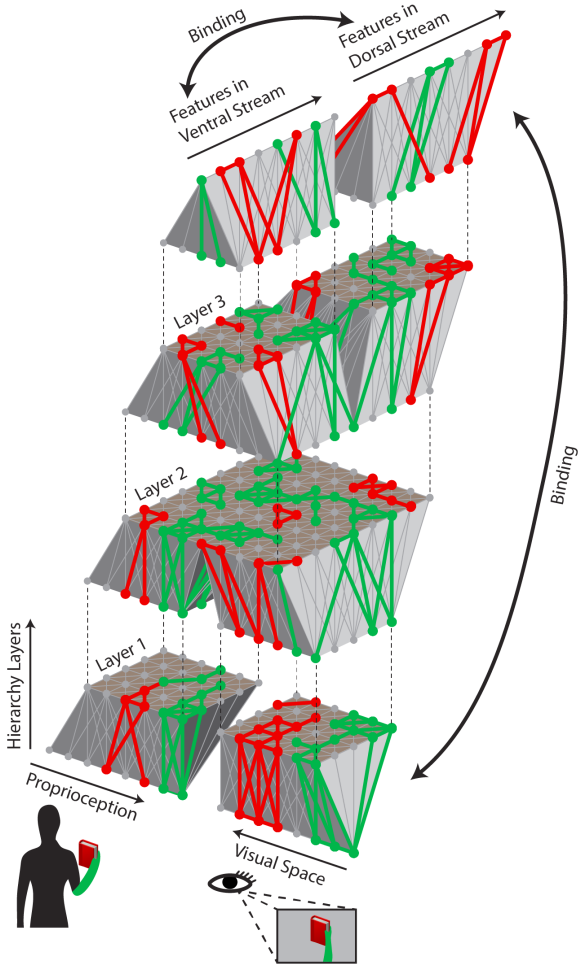


FIGURE 5.4: **Hierarchical model with dynamic coordination.** Exploded-view showing 4 modules (layers) of the hierarchical model. The distributed representations of the hand of the subject and a book are coded by different phases (depicted in green and red). These phase variables allow binding across the two sensory modalities and across the ventral and dorsal streams.

higher level areas. Also, the object representation can be distributed across different higher cortical modules, such that the ventral stream could code the object identity while the dorsal stream could code the position (Chao and Martin, 2000). The distributed representation across cortical modules could be bound together by the phase variables, shown color-coded by green and red edges in the figure. At the same time, two physical objects can be represented simultaneously not only in multiple lower layers but also in multiple higher cortical modules.

Further investigations into the functions of top-down and bottom-up phase interactions would be an interesting area of future research. For example, there is a large body of literature showing that different frequency ranges in the cortex fulfill different roles in coordinating brain functions (for a review see Klimesch, 1999 and Başar et al., 2001). To model the functional role of these different frequencies, it would be interesting to extend the network of coupled phase oscillators (chapter 3) with multiple phase variables each coding the couplings in different frequency ranges. This could be achieved by a Taylor decomposition of the activation variables and a Fourier decomposition of the phase variables (Jirsa and Müller, 2013). Jirsa and Müller used such a formalism to analyze cross frequency couplings in the large-scale connectome.

One can speculate, that a binding of distributed representations within and across modules can significantly increase the flexibility of cognitive functions that such a network can implement. Obviously, the proposed architecture (Figure 5.4) is still missing many higher cognitive and executive control mechanisms. Therefore, an extension with higher specialized modules would be a natural next step that could be discussed. This is especially interesting in regard to recent findings of theta-phase-precession in the hippocampus and the corresponding theories how this could relate to items in working memory (Buzsáki and Moser, 2013). If each item in working memory is activated at different theta-phases in the hippocampus, this activation could directly be transformed into activations with the corresponding phases in the top layer of the proposed architecture.

From a machine learning perspective, there are several network architectures that could implement higher cognitive functions on top of the proposed architecture. For example, the role of the hippocampal memory system could be substituted by a neural Turing machine (Graves, Wayne, and Danihelka, 2014).

These neural Turing machines extend recurrent neural networks with external memory that can be written from network activations and can be read back into network activations. The interface to this memory that controls the locations of the read and write head is controlled by the network itself. Therefore, the control mechanism of this memory interface can be learned by gradient descent optimization with end-to-end back-propagation through the memory matrix. Recently, Graves and colleagues trained such networks to solve reasoning and inference problems in natural language processing and puzzle games (Graves et al., 2016). A more detailed comparison to the role of the hippocampus can also be found in that work. In summary, end to end back-propagation through recurrent networks with an interface to an external memory vastly increases the flexibility and therefore widens the scope of the problems, which can be solved by such an artificial neural networks.

5.4 Conclusion

There are still many remaining questions that are left unanswered and which justify further investigations. What biological mechanisms are really functionally relevant for our cognitive abilities? And which mechanisms can be abstracted away because they are not fulfilling any meaningful cognitive functions? Is there some intermediate level of abstraction that could explain the relevant mechanisms?

In the past, most computational models of neural networks were either biologically realistic simulations reproducing brain functions or artificial neural networks for machine learning applications. The implementation of activation-based neuronal models with phase variables, as presented in this thesis, falls somewhere in between these two categories. Models of this kind could give rise to entirely new opportunities that were so far unexplored in machine learning applications. One can speculate that the more complex the problems become that need to be solved by machine learning methods, the more similar these artificial neural networks become to real biological neural networks.

In this thesis, I have shown that oscillations and synchronization play a crucial role in the coordination of brain activations across and within brain areas. But there are still many other biological mechanisms involved, where the computational principle is not yet understood. Conversely, there are also still many cognitive abilities that cannot be implemented with artificial neural networks. Therefore, this is still an interesting area of further research. Probably, artificial neural networks for machine learning will continue to advance with an enormous pace. Problems that were previously thought to be only solvable by humans, can now be solved with ever more complex deep learning architectures. In this regard, I believe that the path forward is a more integrated view of deep learning and the dynamic organization of brain activations. This could not only advance existing machine learning tools, but could also vastly improve our knowledge about how the brain works.

Appendix A

Connectome Supplementary Material

Empirical Data

Participants

Seventeen healthy volunteers (7 women, mean age 65.6$\pm$10.9 std) underwent DTI and EEG resting-state recordings. None of the participants reported any history of serious medical, neurological or psychiatric diseases. None of the participants were taking any central nervous system-active medication. The study design was approved by the Local Ethical Committee of the Medical Association of Hamburg (PV 3777). All participants gave their written informed consent according to the ethical declaration of Helsinki.

MRI data acquisition

Structural imaging data were acquired using a 3 Tesla Siemens Skyra MRI scanner (Siemens, Erlangen, Germany) and a 32-channel head coil to acquire both diffusion-weighted and high-resolution T1-weighted anatomical images. For diffusion-weighted imaging, 75 axial slices were obtained covering the whole brain with gradients (b=1500 mm^2/s) applied along 64 non-collinear directions with the sequence parameters: Repetition (TR) = 10000 ms, echo time (TE) = 82 ms, field of view (FOV) = 256x204, slice thickness (ST) =

2 mm, in-plane resolution (IPR) = 2x2 mm. The complete dataset consisted of 2 x 64 b1500 images and additionally one b0 image at the beginning and one after the first 64 images. For anatomical imaging, a three-dimensional magnetization-prepared, rapid acquisition gradient-echo sequence (MPRAGE) was used with the following parameters: TR = 2500 ms, TE = 2.12 ms, FOV = 240x192 mm, 256 axial slices, ST = 0.94 mm, IPR = 0.94 x 0.94 mm.

DTI data preprocessing and cortical parcellation

Diffusion-weighted images were analysed using the FSL software package 5.1 (http://www.fmrib.ox.ac.uk/fsl). All datasets were corrected for eddy currents and head motion. Fractional anisotropy (FA) maps were calculated fitting the diffusion tensor model at each voxel. Structural T1-weighted anatomical images were processed using the Freesurfer software package 5.3.0 with standard procedures and parameters resulting in a cortical parcellation of 68 cortical regions Dale, Fischl, and Sereno, 1999; Fischl and Dale, 2000; Fischl, 2012. Accuracy of cortical parcellation was checked visually. Two homologous regions (left and right entorhinal cortex) were discarded from further analysis due to frequent imaging artefacts surrounding this area of the brain. The remaining set of 66 parcellated brain regions were used for further analysis. Registration of structural and diffusion images was achieved using linear and non-linear transformation tools implemented in FSL Jenkinson et al., 2012. Each cortical parcellation was transformed to diffusion space using the non-linear transformation coefficient file and accuracy of registration checked individually.

Fiber tractography and structural connectome construction

Processing of diffusion data included application of a probabilistic diffusion model, modified to allow estimation of multiple (n=2) fiber directions using the program bedpostx Behrens et al., 2007; Behrens et al., 2003. From each seed ROI voxel, 10000 samples were initiated through the probability distribution on principle fiber direction. Tracking resulted in individual maps representing the connectivity value between the seed ROI and individual voxels. Structural connectivity between two regions was measured masking each seed

ROI results by each of the remaining ROI's. In probabilistic tractography, connectivity distribution drops with distance from the seed mask. We calculated the average length between different ROI's using the distance-correction option of probtrackx following recommendations from the online documentation of the FSL library. Values of average distances between seed and target ROI were applied alternatively to the reference method (Figure 1) to account for the confounding effect of tract length.

EEG data acquisition and analysis

Continuous EEG was recorded from 63 cephalic active surface electrodes arranged in the 10/10 system (actiCAP®, Brain Products GmbH, Gilching, Germany) during eight minutes eyes-open resting-state. Impedance was kept below 20 kΩ. Data were sampled at 1000 Hz, referenced to the Cz-electrode (actiCHamp® amplifier, Brain Products GmbH, Gilching). One electrode was mounted below the left eye for EOG-recording. Electrode positions were registered using an ultrasound localization system (CMS20, Zebris, Isny, Germany) before EEG-recording. Subjects were instructed to fixate a stationary fixation cross (viewangle ± 5 to reduce eye movements and were asked to avoid eye blinks, swallowing, any other movements and mental tasks like counting. The continuous EEG was offline rereferenced to a common cephalic average, demeaned, detrended and subjected to an independent component analysis (logistic infomax ICA; Makeig et al., 1996) to remove eye-blink artifacts which were mostly reflected in 1-2 components. The data was downsampled to 125 Hz and segments containing artifacts like muscle activity, lead movements, electrode artifacts or incompletely rejected blink artifacts were removed visually.

The source activity was reconstructed using different inverse solutions:

1. an LCMV beamformer constrained by the covariance of the sensor data Van Veen et al., 1997,
2. an ELORETA spatial filter Pascual-Marqui, 2007a or
3. the MNE Dale et al., 2000.

As a forward model we computed a boundary element method volume conduction model Oostenveld et al., 2003 based on individual T1-weighted structural

MRI of the whole brain and individual electrode positions using the source space modeling functions of the SPM12 toolbox. The source time series were band pass filtered at the alpha frequency band (8±2 Hz) and a Hilbert transform applied. From there functional connectivity estimates were derived as explained above. Analysis was performed with the FieldTrip package for MEG/EEG data analysis Oostenveld et al., 2010 and the Statistical Parametric Mapping software (SPM12b, Wellcome Trust Centre for Neuroimaging, London, UK, http://www.fil.ion.ucl.ac.uk/spm) on MATLAB Version 7.12.0 (R2011a, The Mathworks Inc., Massachusetts, USA).

Figure S1

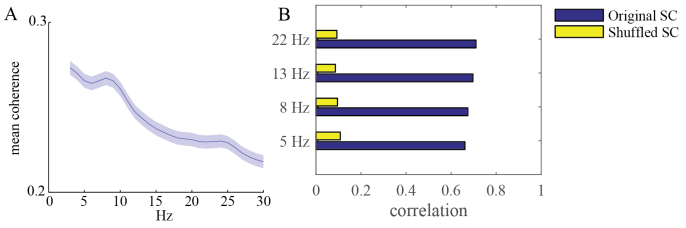


FIGURE A.1: **Evaluation of different EEG frequencies.** A: The mean coherence values (±SEM, shaded area) between all ROIs (n = 2145) is calculated for the frequency range of 3–30 Hz. Overall coherence at lower frequencies is higher with a peak around 8 Hz and a smaller peak around 24 Hz. B: The model performance at different bandpass filters of the EEG source time series. doi:10.1371/journal.pcbi.1005025.s002

Figure S2

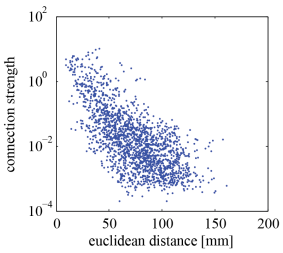


FIGURE A.2: **Dependence between connection strength and euclidean distance.** The euclidean distance is measured between the center coordinates of individual ROIs. The strength between ROIs are the number of tracked DTI fibers divided by the product of both ROI sizes. The logarithm of the structural connection strength is inversely correlated with the euclidean distance (r = -0.37, n = 1883, p < .0001). Connections with zero strength (pairs of ROIs with no probabilistic tracked fibers between them) were excluded (n = 262) due to the logarithmic axis.

doi:10.1371/journal.pcbi.1005025.s003

Appendix B

Behavioral Supplementary Material

Analysis of Variance for Learning Effects

TABLE B.1: **Augmented condition**

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial η^2
Corrected Model	154,286a	5	30,857	,347	,883	,011
Intercept	972192,857	1	972192,857	10945,639	,000	,985
Session	19,000	2	9,500	,107	,899	,001
Half	38,095	1	38,095	,429	,513	,003
Session * Half	97,190	2	48,595	,547	,580	,007
Error	14388,857	162	88,820			
Total	986736,000	168				
Corrected Total	14543,143	167				

TABLE B.2: **Native condition**

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial η^2
Corrected Model	77,643a	5	15,529	,212	,957	,007
Intercept	1093517,357	1	1093517,357	14940,526	,000	,989
Session	32,714	2	16,357	,223	,800	,003
Half	36,214	1	36,214	,495	,483	,003
Session * Half	8,714	2	4,357	,060	,942	,001
Error	11857,000	162	73,191			
Total	1105452,000	168				
Corrected Total	11934,643	167				

TABLE B.3: **Bimodal condition**

Source	**Type III Sum of Squares**	**df**	**Mean Square**	**F**	**Sig.**	**Partial η^2**
Corrected Model	81,262a	5	16,252	,242	,943	,007
Intercept	1056402,881	1	1056402,881	15729,735	,000	,990
Session	60,333	2	30,167	,449	,639	,006
Half	8,595	1	8,595	,128	,721	,001
Session * Half	12,333	2	6,167	,092	,912	,001
Error	10879,857	162	67,160			
Total	1067364,000	168				
Corrected Total	10961,119	167				

Questionnaires

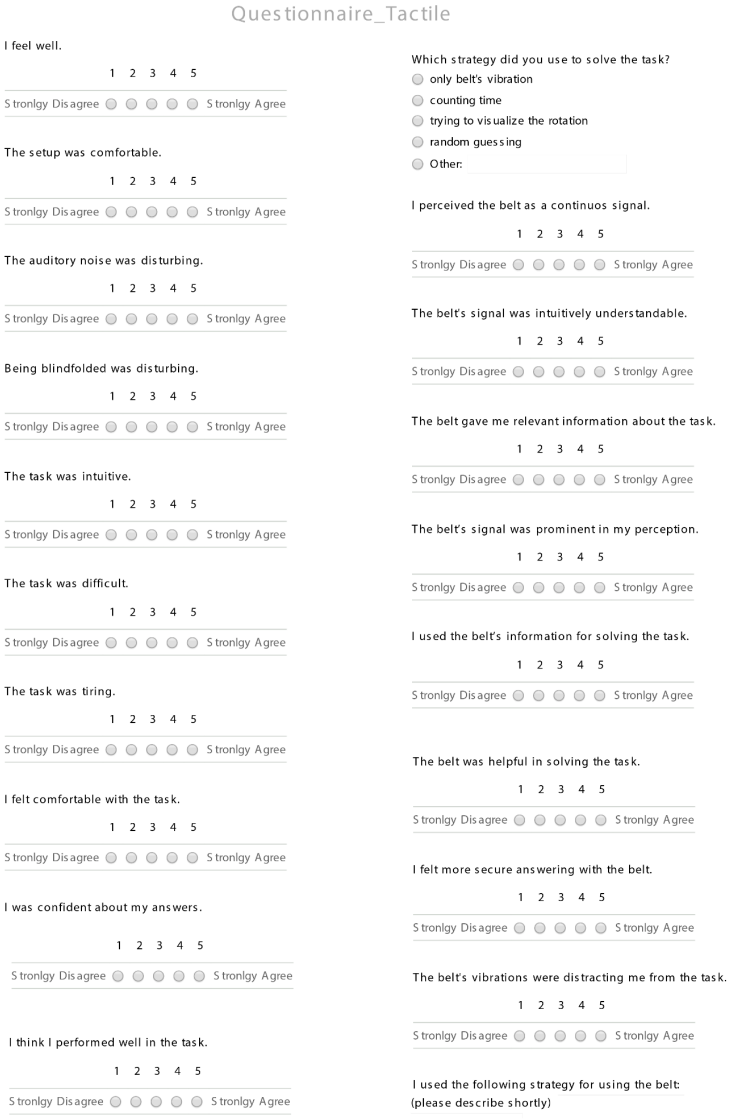


FIGURE B.1: **Questionnaire Augmented Condition.**

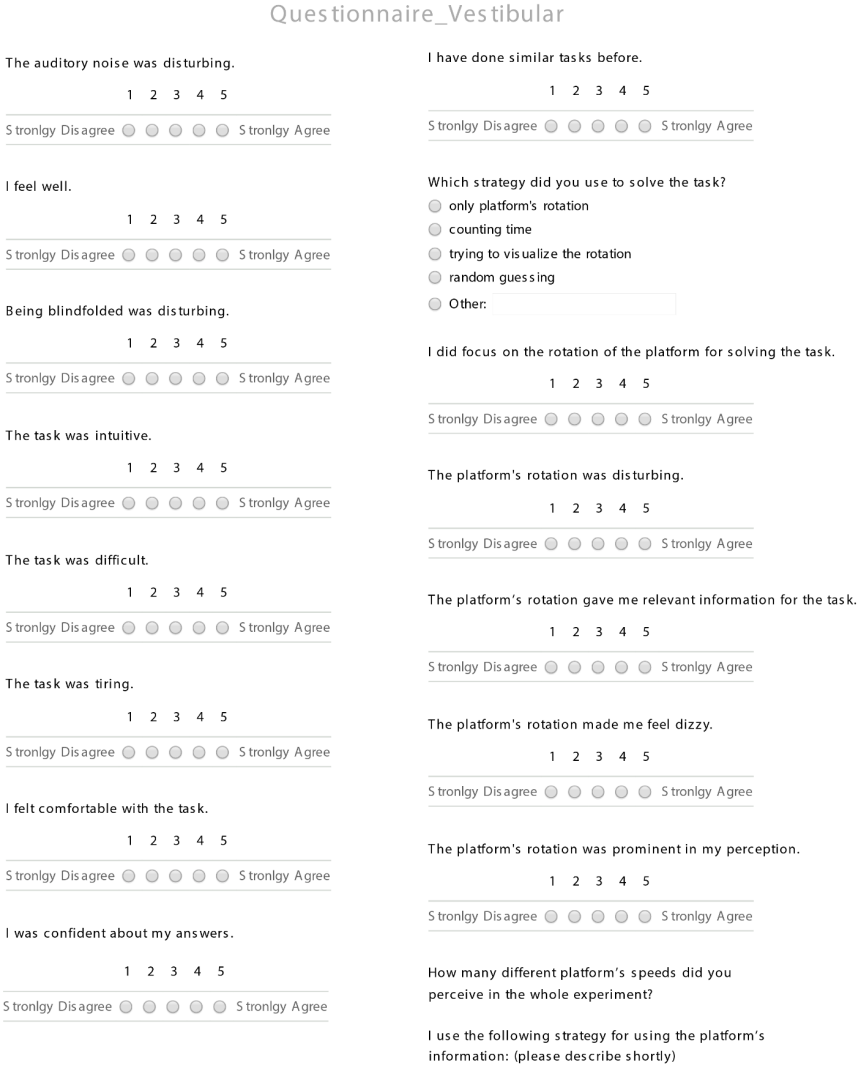


FIGURE B.2: **Questionnaire Native Condition.**

FIGURE B.3: **Questionnaire Bimodal Condition Page 1.**

FIGURE B.4: **Questionnaire Bimodal Condition Page 2.**

Acknowledgments

This thesis would not have been possible without so many people supporting me in so many ways. Firstly, I would like to express my sincere gratitude to my supervisors who have supported me always in the best possible way. Specifically, I would like to thank Professor Peter König for his patience, motivation, and immense knowledge in so many lines of research. Thank you for the continuous guidance and for eliciting my interest in normative models of brain functions and architectures of deep learning, which provided me an eye opening new perspective on analyzing the computational principles of the brain.

I would also like to thank my second supervisor Professor Claus Hilgetag from the University Medical Center Hamburg-Eppendorf. It was a pleasure to spent four weeks in the Institute for Computational Neuroscience in Hamburg. Thank you for the many discussions about my projects that were helpful in so many ways. Your wealth of experience and know-how in the analysis of networks, computational models, brain imaging, and unsupervised learning were often very helpful in advancing my work. After all, I would also like to take this opportunity to thank Professor Christian Gerloff and Professor Andreas Engel for the collaboration within the SFB 936 and the continuous support of my work.

Moreover, I would like to thank all other co-authors of the studies presented in this thesis. I would like to specifically thank my collaborators in Hamburg, Marlene Bönstrup, Bastian Cheng, and Arnaud Messé, for providing me with empirical data and discussions about the related analysis methods. Furthermore, I would like to thank Caspar Goeke for the discussions, that were sometimes intense but always helpful and resulting in the best possible outcomes, and also for proof reading this thesis. In this regard, I would also like to thank Moritz Köster for the close collaboration and conversations about

research related to brain oscillations, multi-site-communication, and memory consolidation, and for proof reading.

I would also like to thank all my other colleagues in the Neurobiopsychology Group and the Institute of Cognitive Science in Osnabrück. Specifically, I would like to thank my former colleague Robert Märtin, who made my start in the lab much easier due to many helpful discussions about artificial neural networks and shared interests in other related fields of research.

Finally, I would also like to thank my family and friends for their support throughout my life.

Declaration

All experiments reported in this thesis conform with the Declaration of Helsinki and have been approved by the ethics committees of the respective institution (University of Osnabrück, Osnabrück, Germany; University Medical Center Hamburg-Eppendorf, Hamburg, Germany). I hereby confirm that I wrote this thesis independently and that I have not made use of resources other than those indicated. I have significantly contributed to all materials used in this thesis. Furthermore, this thesis was neither published in Germany nor abroad, except the parts indicated above, and has not been used to fulfill any other examination requirements.

Bibliography

Abbott, Larry F (1999). "Lapicque's introduction of the integrate-and-fire model neuron (1907)". In: *Brain research bulletin* 50.5, pp. 303–304.

Abeles, Moshe (1982). *Local cortical circuits: An electrophysiological study*. Vol. 6. Springer-Verlag New York.

Açik, A, S Onat, F Schumann, W Einhäuser, and P König (2009). "Effects of luminance contrast and its modifications on fixation behavior during free viewing of images from different categories." In: *Vision research* 49.12, p. 1541.

Adams, Wendy J (2016). "The development of audio-visual integration for temporal judgements". In: *PLoS Comput Biol* 12.4, e1004865.

Aguirre, Geoffrey K, E Zarahn, and M d'Esposito (1997). "Empirical analyses of BOLD fMRI statistics". In: *Neuroimage* 5.3, pp. 199–212.

Akam, Thomas and Dimitri M Kullmann (2010). "Oscillations and filtering networks support flexible routing of information". In: *Neuron* 67.2, pp. 308–320.

Alger, Sara E, Hiuyan Lau, and William Fishbein (2010). "Delayed onset of a daytime nap facilitates retention of declarative memory". In: *PloS one* 5.8, e12131.

Alsius, Agnès, Jordi Navarra, Ruth Campbell, and Salvador Soto-Faraco (2005). "Audiovisual integration of speech falters under high attention demands". In: *Current Biology* 15.9, pp. 839–843.

Alvarez, George A and Patrick Cavanagh (2004). "The capacity of visual short-term memory is set both by visual information load and by number of objects". In: *Psychological science* 15.2, pp. 106–111.

Amit, Daniel J (1992). *Modeling brain function: The world of attractor neural networks*. Cambridge University Press.

Andrew, Colin and Gert Pfurtscheller (1996). "Event-related coherence as a tool for studying dynamic interaction of brain regions". In: *Electroencephalography and clinical neurophysiology* 98.2, pp. 144–148.

Anguera, Joaquin A, Jacqueline Boccanfuso, James L Rintoul, Omar Al-Hashimi, Farhoud Faraji, Jacqueline Janowich, Eric Kong, Yudy Larraburo, Christine Rolle, E Johnston, et al. (2013). "Video game training enhances cognitive control in older adults". In: *Nature* 501.7465, pp. 97–101.

Ao, Xue, Peter Hänggi, and Gerhard Schmid (2013). "In-phase and anti-phase synchronization in noisy Hodgkin–Huxley neurons". In: *Mathematical biosciences* 245.1, pp. 49–55.

Arieli, Amos, Alexander Sterkin, Amiram Grinvald, and Ad Aertsen (1996). "Dynamics of ongoing activity: explanation of the large variability in evoked cortical responses". In: *Science* 273.5283, pp. 1868–1871.

Arnell, Karen M and Julie M Larson (2002). "Cross-modality attentional blinks without preparatory task-set switching". In: *Psychonomic Bulletin & Review* 9.3, pp. 497–506.

Assaf, Yaniv and Ofer Pasternak (2008). "Diffusion tensor imaging (DTI)-based white matter mapping in brain research: a review". In: *Journal of Molecular Neuroscience* 34.1, pp. 51–61.

Bach-Y-Rita, Paul, Carter C Collins, Frank A Saunders, Benjamin White, and Lawrence Scadden (1969). "Vision substitution by tactile image projection". In:

Backus, Alexander R, Jan-Mathijs Schoffelen, Szabolcs Szebényi, Simon Hanslmayr, and Christian F Doeller (2016). "Hippocampal-prefrontal theta oscillations support memory integration". In: *Current Biology* 26.4, pp. 450–457.

Baldi, Pierre (2012). "Autoencoders, unsupervised learning, and deep architectures." In: *ICML unsupervised and transfer learning* 27.37-50, p. 1.

Baldi, Pierre and Kurt Hornik (1989). "Neural networks and principal component analysis: Learning from examples without local minima". In: *Neural networks* 2.1, pp. 53–58.

Baldi, Pierre and Zhiqin Lu (2012). "Complex-valued autoencoders". In: *Neural Networks*.

Barlow, Horace B (1959). "Sensory mechanisms, the reduction of redundancy, and intelligence". In: *The mechanisation of thought processes* 10, pp. 535–539.

Barnett, Lionel, Christopher L Buckley, and Seth Bullock (2009). "Neural complexity and structural connectivity". In: *Physical Review E* 79.5, p. 051914.

Barnett-Cowan, Michael and Laurence R Harris (2009). "Perceived timing of vestibular stimulation relative to touch, light and sound". In: *Experimental brain research* 198.2-3, pp. 221–231.

Başar, Erol, Canan Başar-Eroglu, Sirel Karakaş, and Martin Schürmann (2001). "Gamma, alpha, delta, and theta oscillations govern cognitive processes". In: *International journal of psychophysiology* 39.2, pp. 241–248.

Basser, Peter J, Sinisa Pajevic, Carlo Pierpaoli, Jeffrey Duda, and Akram Aldroubi (2000). "In vivo fiber tractography using DT-MRI data". In: *Magnetic resonance in medicine* 44.4, pp. 625–632.

Bastos, Andre M, Julien Vezoli, and Pascal Fries (2015). "Communication through coherence with inter-areal delays". In: *Current opinion in neurobiology* 31, pp. 173–180.

Bastos, Andre M, W Martin Usrey, Rick A Adams, George R Mangun, Pascal Fries, and Karl J Friston (2012). "Canonical microcircuits for predictive coding". In: *Neuron* 76.4, pp. 695–711.

Bates, Sarah L and Thomas Wolbers (2014). "How cognitive aging affects multisensory integration of navigational cues". In: *Neurobiology of aging* 35.12, pp. 2761–2769.

Battaglia, Peter W, Robert A Jacobs, and Richard N Aslin (2003). "Bayesian integration of visual and auditory signals for spatial localization". In: *JOSA A* 20.7, pp. 1391–1397.

Bauer, Michael, Sean Treichler, Elliott Slaughter, and Alex Aiken (2012). "Legion: expressing locality and independence with logical regions". In: *Proceedings of the international conference on high performance computing, networking, storage and analysis*. IEEE Computer Society Press, p. 66.

Baylor, Denis A (1987). "Photoreceptor signals and vision. Proctor lecture." In: *Investigative ophthalmology & visual science* 28.1, pp. 34–49.

Beggs, John M and Dietmar Plenz (2003). "Neuronal avalanches in neocortical circuits". In: *The Journal of neuroscience* 23.35, pp. 11167–11177.

Behrens, TEJ, H Johansen-Berg, MW Woolrich, SM Smith, CAM Wheeler-Kingshott, PA Boulby, GJ Barker, EL Sillery, K Sheehan, O Ciccarelli, et al. (2003). "Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging". In: *Nature neuroscience* 6.7, pp. 750–757.

Behrens, TEJ, H Johansen Berg, Saad Jbabdi, MFS Rushworth, and MW Woolrich (2007). "Probabilistic diffusion tractography with multiple fibre orientations: What can we gain?" In: *Neuroimage* 34.1, pp. 144–155.

Bell, Anthony J and Terrence J Sejnowski (1997). "The "independent components" of natural scenes are edge filters". In: *Vision research* 37.23, p. 3327.

Belluscio, Mariano A, Kenji Mizuseki, Robert Schmidt, Richard Kempter, and György Buzsáki (2012). "Cross-frequency phase–phase coupling between theta and gamma oscillations in the hippocampus". In: *The Journal of Neuroscience* 32.2, pp. 423–435.

Benchenane, Karim, Adrien Peyrache, Mehdi Khamassi, Patrick L Tierney, Yves Gioanni, Francesco P Battaglia, and Sidney I Wiener (2010). "Coherent theta oscillations and reorganization of spike timing in the hippocampal-prefrontal network upon learning". In: *Neuron* 66.6, pp. 921–936.

Bengio, Yoshua (2009). "Learning deep architectures for AI". In: *Foundations and trends® in Machine Learning* 2.1, pp. 1–127.

Bengio, Yoshua, Aaron Courville, and Pascal Vincent (2013). "Representation learning: A review and new perspectives". In: *IEEE transactions on pattern analysis and machine intelligence* 35.8, pp. 1798–1828.

Bengio, Yoshua, Yann LeCun, et al. (2007). "Scaling learning algorithms towards AI". In: *Large-scale kernel machines* 34.5.

Bengio, Yoshua, Pascal Lamblin, Dan Popovici, Hugo Larochelle, et al. (2007). "Greedy layer-wise training of deep networks". In: *Advances in neural information processing systems* 19, p. 153.

Bengio, Yoshua et al. (2012). "Deep Learning of Representations for Unsupervised and Transfer Learning." In: *ICML Unsupervised and Transfer Learning* 27, pp. 17–36.

Benjamini, Yoav and Daniel Yekutieli (2001). "The control of the false discovery rate in multiple testing under dependency". In: *Annals of statistics*, pp. 1165–1188.

Berger, Hans (1929). "Über das elektrenkephalogramm des menschen". In: *European Archives of Psychiatry and Clinical Neuroscience* 87.1, pp. 527–570.

Berthoz, Alain, Isabelle Israël, Pierre Georges-François, Renato Grasso, and Toshihiro Tsuzuku (1995). "Spatial memory of body linear displacement: what is being stored?" In: *Science* 269.5220, p. 95.

Bertrand, Olivier and Christo Pantev (1994). "Stimulus frequency dependence of the transient oscillatory auditory evoked responses (40 Hz) studied by electric and magnetic recordings in human". In: *Oscillatory event-related brain dynamics*. Springer, pp. 231–242.

Bertrand, Olivier and Catherine Tallon-Baudry (2000). "Oscillatory gamma activity in humans: a possible role for object representation". In: *International Journal of Psychophysiology* 38.3, pp. 211–223.

Betsch, Belinda Y, Wolfgang Einhäuser, Konrad P Körding, and Peter König (2004). "The world from a cat's perspective–statistics of natural videos". In: *Biological cybernetics* 90.1, pp. 41–50.

Betzel, Richard F, Andrea Avena-Koenigsberger, Joaquín Goñi, Ye He, Marcel A de Reus, Alessandra Griffa, Petra E Vértes, Bratislav Mišic, Jean-Philippe Thiran, Patric Hagmann, et al. (2016). "Generative models of the human connectome". In: *Neuroimage* 124, pp. 1054–1064.

Bi, Guo-qiang and Mu-ming Poo (1998). "Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type". In: *The Journal of neuroscience* 18.24, pp. 10464–10472.

Bishop, Christopher M (2006). "Pattern recognition". In: *Machine Learning* 128.

Biswal, Bharat, F Zerrin Yetkin, Victor M Haughton, and James S Hyde (1995). "Functional connectivity in the motor cortex of resting human brain using echo-planar MRI". In: *Magnetic resonance in medicine* 34.4, pp. 537–541.

Blake, Andrew, Heinrich H Bülthoff, and David Sheinberg (1993). "Shape from texture: Ideal observers and human psychophysics". In: *Vision research* 33.12, pp. 1723–1737.

Bojak, Ingo, Thom F Oostendorp, Andrew T Reid, and Rolf Kötter (2010). "Connecting mean field models of neural activity to EEG and fMRI data". In: *Brain topography* 23.2, pp. 139–149.

Born, Jan and Ines Wilhelm (2012). "System consolidation of memory during sleep". In: *Psychological research* 76.2, pp. 192–203.

Breakspear, Michael, Stewart Heitmann, and Andreas Daffertshofer (2010). "Generative models of cortical oscillations: neurobiological implications of the Kuramoto model". In: *Frontiers in human neuroscience* 4.

Breakspear, Michael, Viktor Jirsa, and Gustavo Deco (2010). "Computational models of the brain: from structure to function". In: *Neuroimage* 52.3, pp. 727–730.

Brickenkamp, R (2002). "d2 Aufmerksamkeits-Belastungs-Test [Test of Attention d2; in German]". In: *Gottingen, Germany: Hogrefe*.

Brincat, Scott L and Earl K Miller (2015). "Frequency-specific hippocampal-prefrontal interactions during associative learning". In: *Nature neuroscience* 18.4, pp. 576–581.

Brodmann, Korbinian (1909). *Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues*. Barth.

Brookes, Matthew J, Joanne R Hale, Johanna M Zumer, Claire M Stevenson, Susan T Francis, Gareth R Barnes, Julia P Owen, Peter G Morris, and Srikantan S Nagarajan (2011). "Measuring functional connectivity using MEG: methodology and comparison with fcMRI". In: *Neuroimage* 56.3, pp. 1082–1104.

Brownell, William E, Charles R Bader, Daniel Bertrand, and Yves De Ribaupierre (1985). "Evoked mechanical responses of isolated cochlear outer hair cells". In: *Science* 227.4683, pp. 194–196.

Bullmore, Ed and Olaf Sporns (2009). "Complex brain networks: graph theoretical analysis of structural and functional systems". In: *Nature Reviews Neuroscience* 10.3, pp. 186–198.

– (2012). "The economy of brain network organization". In: *Nature Reviews Neuroscience* 13.5, pp. 336–349.

Buonomano, Dean V and Wolfgang Maass (2009). "State-dependent computations: spatiotemporal processing in cortical networks". In: *Nature Reviews Neuroscience* 10.2, pp. 113–125.

Burbank, Kendra S (2015). "Mirrored STDP Implements Autoencoder Learning in a Network of Spiking Neurons". In: *PLoS Comput Biol* 11.12, e1004566.

Burnett, LR, BE Stein, D Chaponis, and MT Wallace (2004). "Superior colliculus lesions preferentially disrupt multisensory orientation". In: *Neuroscience* 124.3, pp. 535–547.

Burnett, Luke R, Barry E Stein, Thomas J Perrault Jr, and Mark T Wallace (2007). "Excitotoxic lesions of the superior colliculus preferentially impact multisensory neurons and multisensory integration". In: *Experimental brain research* 179.2, pp. 325–338.

Burr, David and Monica Gori (2012). "Multisensory integration develops late in humans". In:

Busch, Niko A, Christoph S Herrmann, Matthias M Müller, Daniel Lenz, and Thomas Gruber (2006). "A cross-laboratory study of event-related gamma activity in a standard object recognition paradigm". In: *Neuroimage* 33.4, pp. 1169–1177.

Butcher, John Charles (1996). "A history of Runge-Kutta methods". In: *Applied numerical mathematics* 20.3, pp. 247–260.

Butler, John S, Stuart T Smith, Jennifer L Campos, and Heinrich H Bülthoff (2010). "Bayesian integration of visual and vestibular signals for heading". In: *Journal of vision* 10.11, pp. 23–23.

Buzsáki, György (1989). “Two-stage model of memory trace formation: a role for “noisy” brain states”. In: *Neuroscience* 31.3, pp. 551–570.

– (1996). “The hippocampo-neocortical dialogue”. In: *Cerebral cortex* 6.2, pp. 81–92.

– (2002). “Theta oscillations in the hippocampus”. In: *Neuron* 33.3, pp. 325–340.

Buzsáki, György and Edvard I Moser (2013). “Memory, navigation and theta rhythm in the hippocampal-entorhinal system”. In: *Nature neuroscience* 16.2, pp. 130–138.

Cabral, Joana, Morten L Kringelbach, and Gustavo Deco (2014). “Exploring the network dynamics underlying brain activity during rest”. In: *Progress in neurobiology* 114, pp. 102–131.

Cabral, Joana, Etienne Hugues, Olaf Sporns, and Gustavo Deco (2011). “Role of local network oscillations in resting-state functional connectivity”. In: *Neuroimage* 57.1, pp. 130–139.

Cabral, Joana, Henry Luckhoo, Mark Woolrich, Morten Joensson, Hamid Mohseni, Adam Baker, Morten L Kringelbach, and Gustavo Deco (2014). “Exploring mechanisms of spontaneous functional connectivity in MEG: how delayed network interactions lead to structured amplitude envelopes of band-pass filtered oscillations”. In: *Neuroimage* 90, pp. 423–435.

Cajal, S Ramón y (1909). “Histologie du système nerveux de l’homme et des vertébrés, Édition Française revue et mise à jour par l’auteur, traduite de l’Espagnol par L. Azoulay”. In:

Calvo, Paco and Toni Gomila (2008). *Handbook of cognitive science: An embodied approach*. Elsevier.

Campo, Adrià Tauste, Marina Martinez-Garcia, Verónica Nácher, Rogelio Luna, Ranulfo Romo, and Gustavo Deco (2015). “Task-driven intra-and interarea communications in primate cerebral cortex”. In: *Proceedings of the National Academy of Sciences* 112.15, pp. 4761–4766.

Carr, Catherine E and Masakazu Konishi (1988). “Axonal delay lines for time measurement in the owl’s brainstem”. In: *Proceedings of the National Academy of Sciences* 85.21, pp. 8311–8315.

Casanova, Manuel F (2005). *Neocortical modularity and the cell minicolumn*. Nova Science Pub Incorporated.

Chao, Linda L and Alex Martin (2000). “Representation of manipulable man-made objects in the dorsal stream”. In: *Neuroimage* 12.4, pp. 478–484.

Chaumon, Maximilien, Dorothy VM Bishop, and Niko A Busch (2015). "A practical guide to the selection of independent components of the electroencephalogram for artifact correction". In: *Journal of neuroscience methods* 250, pp. 47–63.

Chen, Ke and DeLiang Wang (2002). "A dynamically coupled neural oscillator network for image segmentation". In: *Neural Networks* 15.3, pp. 423–439.

Chen, Xiaoli and Timothy P McNamara (2014). "Bayesian Cue Interaction in Human Spatial Navigation". In: *International Conference on Spatial Cognition*. Springer, pp. 147–160.

Cheng, Bastian, Hanna Braass, Christos Ganos, Andras Treszl, Katja Biermann-Ruben, Friedhelm C Hummel, Kirsten Müller-Vahl, Alfons Schnitzler, Christian Gerloff, Alexander Münchau, et al. (2014). "Altered intrahemispheric structural connectivity in Gilles de la Tourette syndrome". In: *NeuroImage: Clinical* 4, pp. 174–181.

Cheng, Hu, Yang Wang, Jinhua Sheng, William G Kronenberger, Vincent P Mathews, Tom A Hummer, and Andrew J Saykin (2012). "Characteristics and variability of structural networks derived from diffusion tensor imaging". In: *Neuroimage* 61.4, pp. 1153–1164.

Chu, CJ, N Tanaka, J Diaz, BL Edlow, O Wu, M Hämäläinen, S Stufflebeam, SS Cash, and MA Kramer (2015). "EEG functional connectivity is partially predicted by underlying white matter connectivity". In: *Neuroimage* 108, pp. 23–33.

Clément, A and Sylvie Droit-Volet (2006). "Counting in a time discrimination task in children and adults". In: *Behavioural Processes* 71.2, pp. 164–171.

Coates, Adam, Honglak Lee, and Andrew Y Ng (2010). "An analysis of single-layer networks in unsupervised feature learning". In: *Ann Arbor* 1001, p. 48109.

Contreras, D and M Steriade (1996). "State-dependent fluctuations of low-frequency rhythms in corticothalamic networks". In: *Neuroscience* 76.1, pp. 25–38.

Cordes, Dietmar, Victor M Haughton, Konstantinos Arfanakis, Gary J Wendt, Patrick A Turski, Chad H Moritz, Michelle A Quigley, and M Elizabeth Meyerand (2000). "Mapping functionally related regions of brain with functional connectivity MR imaging". In: *American Journal of Neuroradiology* 21.9, pp. 1636–1644.

Dale, Anders M, Bruce Fischl, and Martin I Sereno (1999). "Cortical surface-based analysis: I. Segmentation and surface reconstruction". In: *Neuroimage* 9.2, pp. 179–194.

Dale, Anders M, Arthur K Liu, Bruce R Fischl, Randy L Buckner, John W Belliveau, Jeffrey D Lewine, and Eric Halgren (2000). "Dynamic statistical parametric mapping: combining fMRI and MEG for high-resolution imaging of cortical activity". In: *Neuron* 26.1, pp. 55–67.

Dauphin, Yann N, Razvan Pascanu, Caglar Gulcehre, Kyunghyun Cho, Surya Ganguli, and Yoshua Bengio (2014). "Identifying and attacking the saddle point problem in high-dimensional non-convex optimization". In: *Advances in neural information processing systems*, pp. 2933–2941.

David, Olivier and Karl J Friston (2003). "A neural mass model for MEG/EEG:: coupling and neuronal dynamics". In: *NeuroImage* 20.3, pp. 1743–1755.

De Winkel, KN, F Soyka, Michael Barnett-Cowan, HH Bülthoff, EL Groen, and PJ Werkhoven (2013). "Integration of visual and inertial cues in the perception of angular self-motion". In: *Experimental brain research* 231.2, pp. 209–218.

De Winkel, Ksander N, Mikhail Katliar, and Heinrich H Bülthoff (2015). "Forced fusion in multisensory heading estimation". In: *PloS one* 10.5, e0127104.

Deco, Gustavo, Viktor K Jirsa, and Anthony R McIntosh (2011). "Emerging concepts for the dynamical organization of resting-state activity in the brain". In: *Nature Reviews Neuroscience* 12.1, pp. 43–56.

Descoteaux, Maxime, Rachid Deriche, Thomas R Knosche, and Alfred Anwander (2009). "Deterministic and probabilistic tractography based on complex fibre orientation distributions". In: *IEEE transactions on medical imaging* 28.2, pp. 269–286.

Desikan, Rahul S, Florent Ségonne, Bruce Fischl, Brian T Quinn, Bradford C Dickerson, Deborah Blacker, Randy L Buckner, Anders M Dale, R Paul Maguire, Bradley T Hyman, et al. (2006). "An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest". In: *Neuroimage* 31.3, pp. 968–980.

Destexhe, Alain, Diego Contreras, and Mircea Steriade (1999). "Spatiotemporal analysis of local field potentials and unit discharges in cat cerebral cortex during natural wake and sleep states". In: *The Journal of Neuroscience* 19.11, pp. 4595–4608.

Deuker, Lorena, Jan Olligs, Juergen Fell, Thorsten A Kranz, Florian Mormann, Christian Montag, Martin Reuter, Christian E Elger, and Nikolai Axmacher (2013). "Memory consolidation by replay of stimulus-specific neural activity". In: *The Journal of Neuroscience* 33.49, pp. 19373–19383.

Dickerson, Bradford C, I Goncharova, MP Sullivan, C Forchetti, RS Wilson, DA Bennett, Laurel A Beckett, et al. (2001). "MRI-derived entorhinal and hippocampal atrophy in incipient and very mild Alzheimer's disease". In: *Neurobiology of aging* 22.5, pp. 747–754.

Diekelmann, Susanne and Jan Born (2010). "The memory function of sleep". In: *Nature Reviews Neuroscience* 11.2, pp. 114–126.

Donahue, Jeffrey, Lisa Anne Hendricks, Sergio Guadarrama, Marcus Rohrbach, Subhashini Venugopalan, Kate Saenko, and Trevor Darrell (2015). "Long-term recurrent convolutional networks for visual recognition and description". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2625–2634.

Douglas, Rodney J, Kevan AC Martin, and David Whitteridge (1989). "A canonical microcircuit for neocortex". In: *Neural computation* 1.4, pp. 480–488.

Dreiseitl, Stephan and Lucila Ohno-Machado (2002). "Logistic regression and artificial neural network classification models: a methodology review". In: *Journal of biomedical informatics* 35.5, pp. 352–359.

Driver, Jon and Toemme Noesselt (2008). "Multisensory interplay reveals crossmodal influences on 'sensory-specific'brain regions, neural responses, and judgments". In: *Neuron* 57.1, pp. 11–23.

Eichenbaum, Howard (2000). "A cortical–hippocampal system for declarative memory". In: *Nature Reviews Neuroscience* 1.1, pp. 41–50.

Einhaäuser, Wolfgang, Christoph Kayser, Peter Koönig, and Konrad P Koörding (2002). "Learning the invariance properties of complex cells from their responses to natural stimuli". In: *European Journal of Neuroscience* 15.3, pp. 475–486.

Einhäuser, Wolfgang and Peter König (2010). "Getting real—sensory processing of natural stimuli". In: *Current opinion in neurobiology* 20.3, pp. 389–395.

Ellenbogen, Jeffrey M, Jessica D Payne, and Robert Stickgold (2006). "The role of sleep in declarative memory consolidation: passive, permissive, active or none?" In: *Current opinion in neurobiology* 16.6, pp. 716–722.

Elton, Amanda and Wei Gao (2014). "Divergent task-dependent functional connectivity of executive control and salience networks". In: *Cortex* 51, pp. 56–66.

Engel, Andreas K and Pascal Fries (2010). "Beta-band oscillations—signalling the status quo?" In: *Current opinion in neurobiology* 20.2, pp. 156–165.

Engel, Andreas K, Pascal Fries, and Wolf Singer (2001). "Dynamic predictions: oscillations and synchrony in top–down processing". In: *Nature Reviews Neuroscience* 2.10, pp. 704–716.

Engel, Andreas K, Peter König, and Wolf Singer (1991). "Direct physiological evidence for scene segmentation by temporal coding." In: *Proceedings of the National Academy of Sciences* 88.20, pp. 9136–9140.

Engel, Andreas K and Wolf Singer (2001). "Temporal binding and the neural correlates of sensory awareness". In: *Trends in cognitive sciences* 5.1, pp. 16–25.

Engel, Andreas K, Peter König, Charles M Gray, and Wolf Singer (1990). "Stimulus-Dependent Neuronal Oscillations in Cat Visual Cortex: Inter-Columnar Interaction as Determined by Cross-Correlation Analysis". In: *European Journal of Neuroscience* 2.7, pp. 588–606.

Engel, Andreas K, Peter König, Andreas K Kreiter, and Wolf Singer (1991). "Interhemispheric synchronization of oscillatory neuronal responses in cat visual cortex". In: *Science* 252.5009, pp. 1177–1179.

Engel, Andreas K, Peter König, Andreas K Kreiter, Thomas B Schillen, and Wolf Singer (1992). "Temporal coding in the visual cortex: new vistas on integration in the nervous system". In: *Trends in neurosciences* 15.6, pp. 218–226.

Engel, Andreas K, Christian Gerloff, Claus C Hilgetag, and Guido Nolte (2013). "Intrinsic coupling modes: multiscale interactions in ongoing brain activity". In: *Neuron* 80.4, pp. 867–886.

Engel, Stephen A, Gary H Glover, and Brian A Wandell (1997). "Retinotopic organization in human visual cortex and the spatial precision of functional MRI." In: *Cerebral cortex* 7.2, pp. 181–192.

Erhan, Dumitru, Yoshua Bengio, Aaron Courville, Pierre-Antoine Manzagol, Pascal Vincent, and Samy Bengio (2010). "Why does unsupervised pretraining help deep learning?" In: *Journal of Machine Learning Research* 11.Feb, pp. 625–660.

Ernst, Marc O (2007). "Learning to integrate arbitrary signals from vision and touch". In: *Journal of Vision* 7.5, pp. 7–7.

Ernst, Marc O and Martin S Banks (2002). "Humans integrate visual and haptic information in a statistically optimal fashion". In: *Nature* 415.6870, pp. 429–433.

Ernst, Marc O and Heinrich H Bülthoff (2004). "Merging the senses into a robust percept". In: *Trends in cognitive sciences* 8.4, pp. 162–169.

Ernst, MO and G Knoblich (2006). "Human body perception from the inside out". In:

Euston, David R, Masami Tatsuno, and Bruce L McNaughton (2007). "Fast-forward playback of recent memory sequences in prefrontal cortex during sleep". In: *science* 318.5853, pp. 1147–1150.

Falcon, Maria Inez, Jeffrey D Riley, Viktor Jirsa, Anthony R McIntosh, Ahmed D Shereen, E Elinor Chen, and Ana Solodkin (2015). "The Virtual Brain: Modeling Biological Correlates of Recovery after Chronic Stroke". In: *Frontiers in neurology* 6.

Falcon, Maria Inez, Jeffrey D Riley, Viktor Jirsa, Anthony R McIntosh, E Elinor Chen, and Ana Solodkin (2016). "Functional mechanisms of recovery after chronic stroke: modeling with The Virtual Brain". In: *eneuro* 3.2, ENEURO–0158.

Faugloire, Elise and Laure Lejeune (2014). "Evaluation of heading performance with vibrotactile guidance: The benefits of information–movement coupling compared with spatial language." In: *Journal of Experimental Psychology: Applied* 20.4, p. 397.

Fell, Juergen and Nikolai Axmacher (2011). "The role of phase synchronization in memory processes". In: *Nature reviews neuroscience* 12.2, pp. 105–118.

Felleman, Daniel J and David C Van Essen (1991). "Distributed hierarchical processing in the primate cerebral cortex". In: *Cerebral cortex* 1.1, pp. 1–47.

Fetsch, Christopher R, Amanda H Turner, Gregory C DeAngelis, and Dora E Angelaki (2009). "Dynamic reweighting of visual and vestibular cues during self-motion perception". In: *The Journal of Neuroscience* 29.49, pp. 15601–15612.

Fetsch, Christopher R, Alexandre Pouget, Gregory C DeAngelis, and Dora E Angelaki (2012). "Neural correlates of reliability-based cue weighting during multisensory integration". In: *Nature neuroscience* 15.1, pp. 146–154.

Finger, Holger and Peter König (2013). "Phase synchrony facilitates binding and segmentation of natural images in a coupled neural oscillator network". In: *Frontiers in computational neuroscience* 7.195.

Finger, Holger and Shih-Chii Liu (2011). "Estimating the location of a sound source with a spike-timing localization algorithm". In: *2011 IEEE International Symposium of Circuits and Systems (ISCAS)*. IEEE, pp. 2461–2464.

Finger, Holger, Marlene Bönstrup, Bastian Cheng, Arnaud Messé, Claus Hilgetag, Götz Thomalla, Christian Gerloff, and Peter König (2016). "Modeling of Large-Scale Functional Brain Networks Based on Structural Connectivity from DTI: Comparison with EEG Derived Phase Coupling Networks and Evaluation of Alternative Methods along the Modeling Path". In: *PLoS Comput Biol* 12.8, pp. 1–28. DOI: 10.1371/journal.pcbi.1005025. URL: http://dx.doi.org/10.1371%2Fjournal.pcbi.1005025.

Fischl, Bruce (2012). "FreeSurfer". In: *Neuroimage* 62.2, pp. 774–781.

Fischl, Bruce and Anders M Dale (2000). "Measuring the thickness of the human cerebral cortex from magnetic resonance images". In: *Proceedings of the National Academy of Sciences* 97.20, pp. 11050–11055.

Foster, David J and Matthew A Wilson (2006). "Reverse replay of behavioural sequences in hippocampal place cells during the awake state". In: *Nature* 440.7084, pp. 680–683.

Fox, Huw and William Bolton (2002). *Mathematics for engineers and technologists*. Butterworth-Heinemann.

Fox, Michael D, Abraham Z Snyder, Jeffrey M Zacks, and Marcus E Raichle (2006). "Coherent spontaneous activity accounts for trial-to-trial variability in human evoked brain responses". In: *Nature neuroscience* 9.1, pp. 23–25.

Foxe, John J and Adam C Snyder (2011). "The role of alpha-band brain oscillations as a sensory suppression mechanism during selective attention". In: *Frontiers in psychology* 2, p. 154.

Franzius, Mathias, Henning Sprekeler, and Laurenz Wiskott (2007). "Slowness and sparseness lead to place, head-direction, and spatial-view cells". In: *PLoS Computational Biology* 3.8, e166.

Freeman, Linton C (1977). "A set of measures of centrality based on betweenness". In: *Sociometry*, pp. 35–41.

Fries, Pascal (2005). "A mechanism for cognitive dynamics: neuronal communication through neuronal coherence". In: *Trends in cognitive sciences* 9.10, pp. 474–480.

– (2009). "Neuronal gamma-band synchronization as a fundamental process in cortical computation". In: *Annual review of neuroscience* 32, pp. 209–224.

– (2015). "Rhythms for cognition: communication through coherence". In: *Neuron* 88.1, pp. 220–235.

Fries, Pascal, John H Reynolds, Alan E Rorie, and Robert Desimone (2001). "Modulation of oscillatory neuronal synchronization by selective visual attention". In: *Science* 291.5508, pp. 1560–1563.

Friese, Uwe, Moritz Köster, Uwe Hassler, Ulla Martens, Nelson Trujillo Barreto, and Thomas Gruber (2012). "Successful memory encoding is associated with increased cross-frequency coupling between frontal theta and posterior gamma oscillations in human scalp-recorded EEG". In: *Neuroimage*.

Friese, Uwe, Moritz Köster, Uwe Hassler, Ulla Martens, Nelson Trujillo-Barreto, and Thomas Gruber (2013). "Successful memory encoding is associated with increased cross-frequency coupling between frontal theta and posterior gamma oscillations in human scalp-recorded EEG". In: *Neuroimage* 66, pp. 642–647.

Frissen, Ilja, Jennifer L Campos, Jan L Souman, and Marc O Ernst (2011). "Integration of vestibular and proprioceptive signals for spatial updating". In: *Experimental brain research* 212.2, pp. 163–176.

Friston, Karl (2010). "The free-energy principle: a unified brain theory?" In: *Nature Reviews Neuroscience* 11.2, pp. 127–138.

Friston, Karl J (2011). "Functional and effective connectivity: a review". In: *Brain connectivity* 1.1, pp. 13–36.

Friston, Karl J, Lee Harrison, and Will Penny (2003). "Dynamic causal modelling". In: *Neuroimage* 19.4, pp. 1273–1302.

Gais, Steffen, Geneviève Albouy, Mélanie Boly, Thien Thanh Dang-Vu, Annabelle Darsaud, Martin Desseilles, Géraldine Rauchs, Manuel Schabus, Virginie Sterpenich, Gilles Vandewalle, et al. (2007). "Sleep transforms the cerebral trace of declarative memories". In: *Proceedings of the National Academy of Sciences* 104.47, pp. 18778–18783.

Garcés, Pilar, Ernesto Pereda, Juan A Hernández-Tamames, Francisco Del-Pozo, Fernando Maestú, and José Ángel Pineda-Pardo (2015). "Multimodal description of whole brain connectivity: A comparison of resting state MEG, fMRI, and DWI". In: *Human brain mapping*.

Gerloff, Christian, Khalaf Bushara, Alexandra Sailer, Eric M Wassermann, Robert Chen, Takahiro Matsuoka, Daniel Waldvogel, George F Wittenberg, Kenji Ishii, Leonardo G Cohen, et al. (2006). "Multimodal imaging of brain reorganization in motor areas of the contralesional hemisphere of well recovered patients after capsular stroke". In: *Brain* 129.3, pp. 791–808.

Goeke, Caspar M., Serena Planera, Holger Finger, and Peter König (2016). "Bayesian Alternation during Tactile Augmentation". In: *Frontiers in Behavioral Neuroscience* 10, p. 187. ISSN: 1662-5153. DOI: 10.3389/fnbeh.2016.00187. URL: http://journal.frontiersin.org/article/10.3389/fnbeh.2016.00187.

Goldberg, Joshua A, Uri Rokni, Thomas Boraud, Eilon Vaadia, and Hagai Bergman (2004). "Spike synchronization in the cortex-basal ganglia networks of parkinsonian primates reflects global dynamics of the local field potentials". In: *The Journal of neuroscience* 24.26, pp. 6003–6010.

Goñi, Joaquín, Martijn P van den Heuvel, Andrea Avena-Koenigsberger, Nieves Velez de Mendizabal, Richard F Betzel, Alessandra Griffa, Patric Hagmann, Bernat Corominas-Murtra, Jean-Philippe Thiran, and Olaf Sporns (2014). "Resting-brain functional connectivity predicted by analytic measures of network communication". In: *Proceedings of the National Academy of Sciences* 111.2, pp. 833–838.

Goodale, Melvyn A and A David Milner (1992). "Separate visual pathways for perception and action". In: *Trends in neurosciences* 15.1, pp. 20–25.

Gori, Monica, Michela Del Viva, Giulio Sandini, and David C Burr (2008). "Young children do not integrate visual and haptic form information". In: *Current Biology* 18.9, pp. 694–698.

Graves, Alex, Greg Wayne, and Ivo Danihelka (2014). "Neural turing machines". In: *arXiv preprint arXiv:1410.5401*.

Graves, Alex, Greg Wayne, Malcolm Reynolds, Tim Harley, Ivo Danihelka, Agnieszka Grabska-Barwińska, Sergio Gómez Colmenarejo, Edward Grefenstette, Tiago Ramalho, John Agapiou, et al. (2016). "Hybrid computing using a neural network with dynamic external memory". In: *Nature*.

Gray, Charles M (1999). "The temporal correlation hypothesis of visual feature integration: still alive and well". In: *Neuron* 24.1, pp. 31–47.

Gray, Charles M and Wolf Singer (1989). "Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex". In: *Proceedings of the National Academy of Sciences* 86.5, pp. 1698–1702.

Gray, Charles M, Peter König, Andreas K Engel, Wolf Singer, et al. (1989). "Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties". In: *Nature* 338.6213, pp. 334–337.

Gray, Charles M, Andreas K Engel, Peter König, and Wolf Singer (1990). "Stimulus-Dependent Neuronal Oscillations in Cat Visual Cortex: Receptive

Field Properties and Feature Dependence". In: *European Journal of Neuroscience* 2.7, pp. 607–619.

Grech, Roberta, Tracey Cassar, Joseph Muscat, Kenneth P Camilleri, Simon G Fabri, Michalis Zervakis, Petros Xanthopoulos, Vangelis Sakkalis, and Bart Vanrumste (2008). "Review on solving the inverse problem in EEG source analysis". In: *Journal of neuroengineering and rehabilitation* 5.1, p. 1.

Greenblatt, RE, ME Pflieger, and AE Ossadtchi (2012). "Connectivity measures applied to human brain electrophysiological data". In: *Journal of neuroscience methods* 207.1, pp. 1–16.

Gregor, Karol, Ivo Danihelka, Alex Graves, Danilo Jimenez Rezende, and Daan Wierstra (2015). "DRAW: A recurrent neural network for image generation". In: *arXiv preprint arXiv:1502.04623*.

Gregoriou, Georgia G, Stephen J Gotts, Huihui Zhou, and Robert Desimone (2009). "Long-range neural coupling through synchronization with attention". In: *Progress in brain research* 176, pp. 35–45.

Greicius, Michael D, Ben Krasnow, Allan L Reiss, and Vinod Menon (2003). "Functional connectivity in the resting brain: a network analysis of the default mode hypothesis". In: *Proceedings of the National Academy of Sciences* 100.1, pp. 253–258.

Greicius, Michael D, Kaustubh Supekar, Vinod Menon, and Robert F Dougherty (2009). "Resting-state functional connectivity reflects structural connectivity in the default mode network". In: *Cerebral cortex* 19.1, pp. 72–78.

Grondin, Simon, Ginette Meilleur-Wells, and Renée Lachance (1999). "When to start explicit counting in a time-intervals discrimination task: A critical point in the timing process of humans." In: *Journal of Experimental Psychology: Human Perception and Performance* 25.4, p. 993.

Grossman, Emily, M Donnelly, R Price, D Pickens, V Morgan, G Neighbor, and R Blake (2000). "Brain areas involved in perception of biological motion". In: *Journal of cognitive neuroscience* 12.5, pp. 711–720.

Gruber, Thomas and Matthias M Müller (2006). "Oscillatory brain activity in the human EEG during indirect and direct memory tasks". In: *Brain research* 1097.1, pp. 194–204.

Gruber, Thomas, Dimitris Tsivilis, Claire-Marie Giabbiconi, and Matthias M Müller (2008). "Induced electroencephalogram oscillations during source memory: familiarity is reflected in the gamma band, recollection in the theta band". In: *Journal of Cognitive Neuroscience* 20.6, pp. 1043–1053.

Gutkin, Boris, G Bard Ermentrout, and Michael Rudolph (2003). "Spike generating dynamics and the conditions for spike-time precision in cortical neurons". In: *Journal of computational neuroscience* 15.1, pp. 91–103.

Hagmann, Patric, Leila Cammoun, Xavier Gigandet, Reto Meuli, Christopher J Honey, Van J Wedeen, and Olaf Sporns (2008). "Mapping the structural core of human cerebral cortex". In: *PLoS Biol* 6.7, e159.

Hampson, Michelle, Naomi R Driesen, Pawel Skudlarski, John C Gore, and R Todd Constable (2006). "Brain connectivity related to working memory performance". In: *The Journal of neuroscience* 26.51, pp. 13338–13343.

Hanslmayr, Simon, Bernhard Spitzer, and Karl-Heinz Bäuml (2009). "Brain oscillations dissociate between semantic and nonsemantic encoding of episodic memories". In: *Cerebral cortex* 19.7, pp. 1631–1640.

Hassler, Uwe, Nelson Trujillo Barreto, and Thomas Gruber (2011). "Induced gamma band responses in human EEG after the control of miniature saccadic artifacts". In: *Neuroimage* 57.4, pp. 1411–1421.

Hassler, Uwe, Uwe Friese, Ulla Martens, Nelson Trujillo-Barreto, and Thomas Gruber (2013). "Repetition priming effects dissociate between miniature eye movements and induced gamma-band responses in the human electroencephalogram". In: *European Journal of Neuroscience* 38.3, pp. 2425–2433.

He, Biyu J and Marcus E Raichle (2009). "The fMRI signal, slow cortical potential and consciousness". In: *Trends in cognitive sciences* 13.7, pp. 302–309.

He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun (2015). "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification". In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1026–1034.

He, Yong, Zhang J Chen, and Alan C Evans (2007). "Small-world anatomical networks in the human brain revealed by cortical thickness from MRI". In: *Cerebral cortex* 17.10, pp. 2407–2419.

Hebb, D (1949). *0. The Organization of Behavior. New York: John Weley & Sons*.

Helbig, Hannah B and Marc O Ernst (2007). "Optimal integration of shape information from vision and touch". In: *Experimental Brain Research* 179.4, pp. 595–606.

– (2008). "Visual-haptic cue weighting is independent of modality-specific attention". In: *Journal of vision* 8.1, pp. 21–21.

Helmstaedter, Moritz, Kevin L Briggman, and Winfried Denk (2008). "3D structural imaging of the brain with photons and electrons". In: *Current opinion in neurobiology* 18.6, pp. 633–641.

Hermundstad, Ann M, Danielle S Bassett, Kevin S Brown, Elissa M Aminoff, David Clewett, Scott Freeman, Amy Frithsen, Arianne Johnson, Christine M Tipper, Michael B Miller, et al. (2013). "Structural foundations of resting-state and task-based functional connectivity in the human brain". In: *Proceedings of the National Academy of Sciences* 110.15, pp. 6169–6174.

Herrmann, Christoph S, Daniel Strüber, Randolph F Helfrich, and Andreas K Engel (2015). "EEG oscillations: From correlation to causality". In: *International Journal of Psychophysiology* 103, pp. 12–21.

Hindriks, R, M Woolrich, H Luckhoo, M Joensson, H Mohseni, ML Kringelbach, and G Deco (2015). "Role of white-matter pathways in coordinating alpha oscillations in resting visual cortex". In: *NeuroImage* 106, pp. 328–339.

Hindriks, Rikkert, Michel JAM van Putten, and Gustavo Deco (2014). "Intracortical propagation of EEG alpha oscillations". In: *Neuroimage* 103, pp. 444–453.

Hinton, Geoffrey (2010). "A practical guide to training restricted Boltzmann machines". In: *Momentum* 9, p. 1.

Hinton, Geoffrey E (2007). "Learning multiple layers of representation". In: *Trends in cognitive sciences* 11.10, pp. 428–434.

Hinton, Geoffrey E and Ruslan R Salakhutdinov (2006). "Reducing the dimensionality of data with neural networks". In: *Science* 313.5786, pp. 504–507.

Hinton, Geoffrey E and Terrence Joseph Sejnowski (1999). *Unsupervised learning: foundations of neural computation*. MIT press.

Hinton, Geoffrey E and Richard S Zemel (1994). "Autoencoders, minimum description length, and Helmholtz free energy". In: *Advances in neural information processing systems*, pp. 3–3.

Hipp, Joerg F, Andreas K Engel, and Markus Siegel (2011). "Oscillatory synchronization in large-scale cortical networks predicts perception". In: *Neuron* 69.2, pp. 387–396.

Hipp, Joerg F, David J Hawellek, Maurizio Corbetta, Markus Siegel, and Andreas K Engel (2012). "Large-scale cortical correlation structure of spontaneous oscillatory activity". In: *Nature neuroscience* 15.6, pp. 884–890.

Hirsch, Helmut VB and DN Spinelli (1971). "Modification of the distribution of receptive field orientation in cats by selective visual exposure during development". In: *Experimental brain research* 12.5, pp. 509–527.

Hodgkin, Alan L and Andrew F Huxley (1952). "A quantitative description of membrane current and its application to conduction and excitation in nerve". In: *The Journal of physiology* 117.4, p. 500.

Honey, CJ, O Sporns, Leila Cammoun, Xavier Gigandet, Jean-Philippe Thiran, Reto Meuli, and Patric Hagmann (2009). "Predicting human resting-state functional connectivity from structural connectivity". In: *Proceedings of the National Academy of Sciences* 106.6, pp. 2035–2040.

Horwitz, Barry (1990). "Simulating functional interactions in the brain: a model for examining correlations between regional cerebral metabolic rates". In: *International journal of bio-medical computing* 26.3, pp. 149–170.

– (2003). "The elusive concept of brain connectivity". In: *Neuroimage* 19.2, pp. 466–470.

Hubel, David H and Torsten N Wiesel (1962). "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex". In: *The Journal of physiology* 160.1, pp. 106–154.

Hummel, Friedhelm and Christian Gerloff (2005). "Larger interregional synchrony is associated with greater behavioral success in a complex sensory integration task in humans". In: *Cerebral Cortex* 15.5, pp. 670–678.

Humphries, Mark D, Kevin Gurney, and Tony J Prescott (2006). "The brainstem reticular formation is a small-world, not scale-free, network". In: *Proceedings of the Royal Society of London B: Biological Sciences* 273.1585, pp. 503–511.

Hutcheon, Bruce and Yosef Yarom (2000). "Resonance, oscillation and the intrinsic frequency preferences of neurons". In: *Trends in neurosciences* 23.5, pp. 216–222.

Hyvärinen, Aapo and Patrik Hoyer (2000). "Emergence of phase-and shift-invariant features by decomposition of natural images into independent feature subspaces". In: *Neural computation* 12.7, pp. 1705–1720.

Hyvärinen, Aapo and Patrik O Hoyer (2001). "A two-layer sparse coding model learns simple and complex cell receptive fields and topography from natural images". In: *Vision research* 41.18, pp. 2413–2423.

Iber, Conrad (2007). *The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications*. American Academy of Sleep Medicine.

Inouye, Tatsuji (1909). *Die Sehstörungen bei Schussverletzungen der kortikalen Sehsphäre: nach Beobachtungen an Verwundeten der letzten japanischen Kriege*. W. Engelmann.

Izhikevich, Eugene M and Gerald M Edelman (2008). "Large-scale model of mammalian thalamocortical systems". In: *Proceedings of the national academy of sciences* 105.9, pp. 3593–3598.

Izhikevich, Eugene M et al. (2003). "Simple model of spiking neurons". In: *IEEE Transactions on neural networks* 14.6, pp. 1569–1572.

Jansen, Ben H and Vincent G Rit (1995). "Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns". In: *Biological cybernetics* 73.4, pp. 357–366.

Jbabdi, Saad and Heidi Johansen-Berg (2011). "Tractography: where do we go from here?" In: *Brain connectivity* 1.3, pp. 169–183.

Jeffress, Lloyd A (1948). "A place theory of sound localization." In: *Journal of comparative and physiological psychology* 41.1, p. 35.

Jenkinson, Mark, Christian F Beckmann, Timothy EJ Behrens, Mark W Woolrich, and Stephen M Smith (2012). "Fsl". In: *Neuroimage* 62.2, pp. 782–790.

Jensen, Ole, Jochen Kaiser, and Jean-Philippe Lachaux (2007). "Human gamma-frequency oscillations associated with attention and memory". In: *Trends in neurosciences* 30.7, pp. 317–324.

Jensen, Ole and John E Lisman (2005). "Hippocampal sequence-encoding driven by a cortical multi-item working memory buffer". In: *Trends in neurosciences* 28.2, pp. 67–72.

Jensen, Ole and Ali Mazaheri (2010). "Shaping functional architecture by oscillatory alpha activity: gating by inhibition". In: *Frontiers in human neuroscience* 4, p. 186.

Ji, Daoyun and Matthew A Wilson (2007). "Coordinated memory replay in the visual cortex and hippocampus during sleep". In: *Nature neuroscience* 10.1, pp. 100–107.

Jirsa, Viktor and Viktor Müller (2013). "Cross-frequency coupling in real and virtual brain networks". In: *Frontiers in computational neuroscience* 7, p. 78.

Jirsa, Viktor, Olaf Sporns, Michael Breakspear, Gustavo Deco, and Anthony Randal McIntosh (2010). "Towards the virtual brain: network modeling of the intact and the damaged brain". In: *Archives italiennes de biologie* 148.3, pp. 189–205.

Jirsa, VK, T Proix, D Perdikis, MM Woodman, H Wang, J Gonzalez-Martinez, C Bernard, C Bénar, M Guye, P Chauvel, et al. (2016). "The virtual epileptic patient: individualized whole-brain models of epilepsy spread". In: *NeuroImage*.

Jolicoeur, Pierre (1999). "Restricted attentional capacity between sensory modalities". In: *Psychonomic Bulletin & Review* 6.1, pp. 87–92.

Jonas, Josr B, Andreas M Schmidt, JA Müller-Bergh, UM Schlötzer-Schrehardt, and GO Naumann (1992). "Human optic nerve fiber count and optic disc size." In: *Investigative ophthalmology & visual science* 33.6, pp. 2012–2018.

Jones, Matthew W and Matthew A Wilson (2005). "Theta rhythms coordinate hippocampal–prefrontal interactions in a spatial memory task". In: *PLoS Biol* 3.12, e402.

Junghöfer, Markus, Thomas Elbert, Don M Tucker, and Brigitte Rockstroh (2000). "Statistical control of artifacts in dense array EEG/MEG studies". In: *Psychophysiology* 37.04, pp. 523–532.

Kaliuzhna, Mariia, Mario Prsa, Steven Gale, Stella J Lee, and Olaf Blanke (2015). "Learning to integrate contradictory multisensory self-motion cue pairings". In: *Journal of vision* 15.1, pp. 10–10.

Kaplan, Raphael, Daniel Bush, Mathilde Bonnefond, Peter A Bandettini, Gareth R Barnes, Christian F Doeller, and Neil Burgess (2014). "Medial prefrontal theta phase coupling during spatial memory retrieval". In: *Hippocampus* 24.6, pp. 656–665.

Kärcher, Silke Manuela, Sandra Fenzlaff, Daniela Hartmann, Saskia Kathi Nagel, and Peter König (2012). "Sensory augmentation for the blind". In: *Frontiers in human neuroscience* 6, p. 37.

Kaspar, Kai, Sabine König, Jessika Schwandt, and Peter König (2014). "The experience of new sensorimotor contingencies by sensory augmentation". In: *Consciousness and cognition* 28, pp. 47–63.

Kawamoto, Alan H and James A Anderson (1985). "A neural network model of multistable perception". In: *Acta Psychologica* 59.1, pp. 35–65.

Kay, Jim, Dario Floreano, and William A Phillips (1998). "Contextually guided unsupervised learning using local multivariate binary processors". In: *Neural Networks* 11.1, pp. 117–140.

Kay, Jim and William A Phillips (1997). "Activation functions, computational goals, and learning rules for local processors with contextual guidance". In: *Neural Computation* 9.4, pp. 895–910.

Kay, Jim W and WA Phillips (2011). "Coherent infomax as a computational goal for neural systems". In: *Bulletin of mathematical biology* 73.2, pp. 344–372.

Kayser, Christoph, Wolfgang Einhäuser, and Peter König (2003). "Temporal correlations of orientations in natural scenes". In: *Neurocomputing* 52, pp. 117–123.

Kelso, JA Scott (2012). "Multistability and metastability: understanding dynamic coordination in the brain". In: *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 367.1591, pp. 906–918.

Khadem, Ali and Gholam-Ali Hossein-Zadeh (2014). "Quantification of the effects of volume conduction on the EEG/MEG connectivity estimates: an index of sensitivity to brain interactions". In: *Physiological measurement* 35.10, p. 2149.

Kim, Seunghwan, Seon Hee Park, and CS Ryu (1997). "Multistability in coupled oscillator systems with time delay". In: *Physical Review Letters* 79.15, p. 2911.

Kingma, Diederik and Jimmy Ba (2014). "Adam: A method for stochastic optimization". In: *arXiv preprint arXiv:1412.6980*.

Kinoshita, Manabu, Kei Yamada, Naoya Hashimoto, Amami Kato, Shuichi Izumoto, Takahito Baba, Motohiko Maruno, Tsunehiko Nishimura, and Toshiki Yoshimine (2005). "Fiber-tracking does not accurately estimate size of fiber bundle in pathological condition: initial neurosurgical experience using neuronavigation and subcortical white matter stimulation". In: *Neuroimage* 25.2, pp. 424–429.

Klemen, Jane, C Büchel, and M Rose (2009). "Perceptual load interacts with stimulus processing across sensory modalities". In: *European Journal of Neuroscience* 29.12, pp. 2426–2434.

Klimesch, Wolfgang (1997). "EEG-alpha rhythms and memory processes". In: *International Journal of Psychophysiology* 26.1, pp. 319–340.

– (1999). "EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis". In: *Brain research reviews* 29.2, pp. 169–195.

– (2012). "Alpha-band oscillations, attention, and controlled access to stored information". In: *Trends in cognitive sciences* 16.12, pp. 606–617.

Klimesch, Wolfgang, Paul Sauseng, and Simon Hanslmayr (2007). "EEG alpha oscillations: the inhibition–timing hypothesis". In: *Brain research reviews* 53.1, pp. 63–88.

Knill, David C and Jeffrey A Saunders (2003). "Do humans optimally integrate stereo and texture information for judgments of surface slant?" In: *Vision research* 43.24, pp. 2539–2558.

Ko, Andrew L, Felix Darvas, Andrew Poliakov, Jeffrey Ojemann, and Larry B Sorensen (2011). "Quasi-periodic fluctuations in default mode network electrophysiology". In: *The Journal of Neuroscience* 31.32, pp. 11728–11732.

König, Peter, Andreas K Engel, Siegrid Löwel, and Wolf Singer (1993). "Squint affects synchronization of oscillatory responses in cat visual cortex". In: *European Journal of Neuroscience* 5.5, pp. 501–508.

Körding, Konrad P and Daniel M Wolpert (2004). "Bayesian integration in sensorimotor learning". In: *Nature* 427.6971, pp. 244–247.

– (2006). "Bayesian decision theory in sensorimotor control". In: *Trends in cognitive sciences* 10.7, pp. 319–326.

Körding, Konrad P, Christoph Kayser, Wolfgang Einhäuser, and Peter König (2004). "How are complex cell properties adapted to the statistics of natural stimuli?" In: *Journal of Neurophysiology* 91.1, pp. 206–212.

Körding, Konrad P, Ulrik Beierholm, Wei Ji Ma, Steven Quartz, Joshua B Tenenbaum, and Ladan Shams (2007). "Causal inference in multisensory perception". In: *PLoS one* 2.9, e943.

Köster, Moritz, Uwe Friese, Benjamin Schöne, Nelson Trujillo-Barreto, and Thomas Gruber (2014). "Theta–gamma coupling during episodic retrieval in the human EEG". In: *Brain research* 1577, pp. 57–68.

Köster, Moritz, Holger Finger, Maren Kater, Christoph Schenk, and Thomas Gruber (2016). "Neuronal Oscillations Indicate Sleep-dependent Changes in the Cortical Memory Trace". In: *Journal of Cognitive Neuroscience*.

Krusienski, Dean J, Dennis J McFarland, and Jonathan R Wolpaw (2012). "Value of amplitude, phase, and coherence features for a sensorimotor rhythm-based brain–computer interface". In: *Brain research bulletin* 87.1, pp. 130–134.

Kuang, Shenbing and Tao Zhang (2014). "Smelling directions: Olfaction modulates ambiguous visual motion perception". In: *Scientific reports* 4.

Kuhn, Marion, Elias Wolf, Jonathan G Maier, Florian Mainberger, Bernd Feige, Hanna Schmid, Jan Bürklin, Sarah Maywald, Volker Mall, Nikolai H Jung, et al. (2016). "Sleep recalibrates homeostatic and associative synaptic plasticity in the human cortex". In: *Nature Communications* 7.

Kumar, Arvind, Stefan Rotter, and Ad Aertsen (2010). "Spiking activity propagation in neuronal networks: reconciling different perspectives on neural coding". In: *Nature Reviews Neuroscience* 11.9, pp. 615–627.

Kuramoto, Yoshiki (1975). "Self-entrainment of a population of coupled nonlinear oscillators". In: *International symposium on mathematical problems in theoretical physics*. Springer, pp. 420–422.

– (1984). *Chemical oscillations, waves, and turbulence*. Vol. 19. Springer.

Lachaux, Jean-Philippe, Eugenio Rodriguez, Jacques Martinerie, Francisco J Varela, et al. (1999). "Measuring phase synchrony in brain signals". In: *Human brain mapping* 8.4, pp. 194–208.

Lakatos, Peter, George Karmos, Ashesh D Mehta, Istvan Ulbert, and Charles E Schroeder (2008). "Entrainment of neuronal oscillations as a mechanism of attentional selection". In: *science* 320.5872, pp. 110–113.

Landsness, EC, F Ferrarelli, S Sarasso, MR Goldstein, BA Riedner, C Cirelli, B Perfetti, C Moisello, MF Ghilardi, and G Tononi (2011). "Electrophysiological traces of visuomotor learning and their renormalization after sleep". In: *Clinical Neurophysiology* 122.12, pp. 2418–2425.

Landy, MS, MS Banks, and DC Knill (2011). *Book of Sensory Cue Integration*.

Lazar, Andreea, Gordon Pipa, and Jochen Triesch (2009). "SORN: a self-organizing recurrent neural network". In: *Frontiers in Computational Neuroscience* 3.

Le, Quoc V, Alexandre Karpenko, Jiquan Ngiam, and Andrew Y Ng (2011). "Ica with reconstruction cost for efficient overcomplete feature learning". In: *Advances in Neural Information Processing Systems* 24, pp. 1017–1025.

LeCun, Yann and Yoshua Bengio (1995). "Convolutional networks for images, speech, and time series". In: *The handbook of brain theory and neural networks* 3361.10, p. 1995.

LeCun, Yann, Bernhard Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne Hubbard, and Lawrence D Jackel (1989). "Backpropagation applied to handwritten zip code recognition". In: *Neural computation* 1.4, pp. 541–551.

LeCun, Yann, Léon Bottou, Yoshua Bengio, and Patrick Haffner (1998). "Gradient-based learning applied to document recognition". In: *Proceedings of the IEEE* 86.11, pp. 2278–2324.

Lee, Honglak, Roger Grosse, Rajesh Ranganath, and Andrew Y Ng (2009). "Convolutional deep belief networks for scalable unsupervised learning of

hierarchical representations". In: *Proceedings of the 26th annual international conference on machine learning*. ACM, pp. 609–616.

Leon, Paula Sanz, Stuart A Knock, M Marmaduke Woodman, Lia Domide, Jochen Mersmann, Anthony R McIntosh, and Viktor Jirsa (2013). "The Virtual Brain: a simulator of primate brain network dynamics". In: *Frontiers in neuroinformatics* 7.

Li, Chunguang and Yuke Li (2011). "Fast and robust image segmentation by small-world neural oscillator networks". In: *Cognitive neurodynamics* 5.2, pp. 209–220.

Li, Longchuan, James K Rilling, Todd M Preuss, Matthew F Glasser, Frederick W Damen, and Xiaoping Hu (2012a). "Quantitative assessment of a framework for creating anatomical brain networks via global tractography". In: *NeuroImage* 61.4, pp. 1017–1030.

Li, Longchuan, James K Rilling, Todd M Preuss, Matthew F Glasser, and Xiaoping Hu (2012b). "The effects of connection reconstruction method on the interregional connectivity of brain networks via diffusion tractography". In: *Human brain mapping* 33.8, pp. 1894–1913.

Lindeman, Robert W, John L Sibert, Erick Mendez-Mendez, Sachin Patil, and Daniel Phifer (2005). "Effectiveness of directional vibrotactile cuing on a building-clearing task". In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, pp. 271–280.

Linsker, Ralph (1988). "Self-organization in a perceptual network". In: *Computer* 21.3, pp. 105–117.

Lisman, John (2005). "The theta/gamma discrete phase code occuring during the hippocampal phase precession may be a more general brain coding scheme". In: *Hippocampus* 15.7, pp. 913–922.

Lisman, John E and Marco AP Idiart (1995). "Storage of 7 plus/minus 2 short-term memories in oscillatory subcycles". In: *Science* 267.5203, p. 1512.

Lisman, John E and Ole Jensen (2013). "The theta-gamma neural code". In: *Neuron* 77.6, pp. 1002–1016.

Liu, Dong C and Jorge Nocedal (1989). "On the limited memory BFGS method for large scale optimization". In: *Mathematical programming* 45.1-3, pp. 503–528.

Liu, Zhong-Wu, Ugo Faraguna, Chiara Cirelli, Giulio Tononi, and Xiao-Bing Gao (2010). "Direct evidence for wake-related increases and sleep-related decreases in synaptic strength in rodent cortex". In: *The Journal of Neuroscience* 30.25, pp. 8671–8675.

Louie, Kenway and Matthew A Wilson (2001). "Temporally structured replay of awake hippocampal ensemble activity during rapid eye movement sleep". In: *Neuron* 29.1, pp. 145–156.

Löwel, S and W Singer (1992). "Selection of intrinsic horizontal connections in the visual cortex by correlated neuronal activity." In: *Science (New York, NY)* 255.5041, p. 209.

Lu, Hongtao (2002). "Chaotic attractors in delayed neural networks". In: *Physics Letters A* 298.2, pp. 109–116.

Lu, Ju (2011). "Neuronal tracing for connectomic studies". In: *Neuroinformatics* 9.2-3, pp. 159–166.

Mainen, Zachary F and Terrence J Sejnowski (1995). "Reliability of spike timing in neocortical neurons". In: *Science* 268.5216, p. 1503.

Makeig, Scott, Anthony J Bell, Tzyy-Ping Jung, Terrence J Sejnowski, et al. (1996). "Independent component analysis of electroencephalographic data". In: *Advances in neural information processing systems*, pp. 145–151.

Malsburg, Ch Von der and Werner Schneider (1986). "A neural cocktail-party processor". In: *Biological cybernetics* 54.1, pp. 29–40.

Malsburg, Christoph Von der (1981). "The correlation theory of brain function". In: *Models of neural networks II: Temporal aspects of coding and information processing in biological systems*, pp. 95–119.

– (1994). "The correlation theory of brain function". In: *Models of neural networks*. Springer, pp. 95–119.

– (1995). "Binding in models of perception and brain function". In: *Current opinion in neurobiology* 5.4, pp. 520–526.

Malsburg, Christoph Von der, William A Phillips, and Wolf Singer (2010). *Dynamic coordination in the brain: from neurons to mind*. MIT Press.

Manganotti, P, C Gerloff, C Toro, H Katsuta, N Sadato, Pet al Zhuang, L Leocani, and M Hallett (1998). "Task-related coherence and task-related spectral power changes during sequential finger movements". In: *Electroencephalography and Clinical Neurophysiology/Electromyography and Motor Control* 109.1, pp. 50–62.

Mantini, Dante, Mauro Gianni Perrucci, Cosimo Del Gratta, Gian Luca Romani, and Maurizio Corbetta (2007). "Electrophysiological signatures of resting state networks in the human brain". In: *Proceedings of the National Academy of Sciences* 104.32, pp. 13170–13175.

Mariño, Jorge, James Schummers, David C Lyon, Lars Schwabe, Oliver Beck, Peter Wiesing, Klaus Obermayer, and Mriganka Sur (2005). "Invariant computations in local cortical networks with balanced excitation and inhibition". In: *Nature neuroscience* 8.2, pp. 194–201.

Maris, Eric, Pascal Fries, and Freek van Ede (2016). "Diverse phase relations among neuronal rhythms and their potential function". In: *Trends in neurosciences* 39.2, pp. 86–99.

Maris, Eric and Robert Oostenveld (2007). "Nonparametric statistical testing of EEG-and MEG-data". In: *Journal of neuroscience methods* 164.1, pp. 177–190.

Markram, Henry, Joachim Lübke, Michael Frotscher, and Bert Sakmann (1997). "Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs". In: *Science* 275.5297, pp. 213–215.

Marr, David and Ellen Hildreth (1980). "Theory of edge detection". In: *Proceedings of the Royal Society of London B: Biological Sciences* 207.1167, pp. 187–217.

Marshall, Lisa and Jan Born (2007). "The contribution of sleep to hippocampus-dependent memory consolidation". In: *Trends in cognitive sciences* 11.10, pp. 442–450.

Marshall, Lisa, Halla Helgadóttir, Matthias Mölle, and Jan Born (2006). "Boosting slow oscillations during sleep potentiates memory". In: *Nature* 444.7119, pp. 610–613.

Martens, Sander, Manasa Kandula, and John Duncan (2010). "Restricted attentional capacity within but not between sensory modalities: an individual differences approach". In: *PLoS One* 5.12, e15280.

Martin, KA and D Whitteridge (1984). "Form, function and intracortical projections of spiny neurones in the striate visual cortex of the cat." In: *The Journal of physiology* 353, p. 463.

Masquelier, Timothee, Thomas Serre, Simon Thorpe, and Tomaso Poggio (2007). *Learning complex cell invariance from natural videos: A plausibility proof*. Tech. rep. DTIC Document.

Maye, Alexander and Markus Werning (2007). "Neuronal synchronization: from dynamic feature binding to object representations". In: *Chaos Complex. Lett* 2, pp. 315–325.

Mazoyer, Bernard, L Zago, E Mellet, S Bricogne, O Etard, O Houde, F Crivello, M Joliot, L Petit, and Nathalie Tzourio-Mazoyer (2001). "Cortical networks for working memory and executive functions sustain the conscious resting state in man". In: *Brain research bulletin* 54.3, pp. 287–298.

McGaugh, James L (1966). "Time-dependent processes in memory storage". In: *Science* 153.3742, pp. 1351–1358.

Mclntosh, AR and F Gonzalez-Lima (1994). "Structural equation modeling and its application to network analysis in functional brain imaging". In: *Human Brain Mapping* 2.1-2, pp. 2–22.

McNaughton, BL, Carol A Barnes, and J O'keefe (1983). "The contributions of position, direction, and velocity to single unit activity in the hippocampus of freely-moving rats". In: *Experimental Brain Research* 52.1, pp. 41–49.

Messe, Arnaud, Habib Benali, and Guillaume Marrelec (2015). "Relating structural and functional connectivity in MRI: A simple model for a complex brain". In: *Medical Imaging, IEEE Transactions on* 34.1, pp. 27–37.

Messé, Arnaud, David Rudrauf, Habib Benali, and Guillaume Marrelec (2014). "Relating structure and function in the human brain: relative contributions of anatomy, stationary dynamics, and non-stationarities". In: *PLoS computational biology* 10.3, e1003530.

Messé, Arnaud, David Rudrauf, Alain Giron, and Guillaume Marrelec (2015). "Predicting functional connectivity from structural connectivity via computational models using MRI: an extensive comparison study". In: *NeuroImage* 111, pp. 65–75.

Meyer-Bäse, Anke, Frank Ohl, and Henning Scheich (1996). "Singular perturbation analysis of competitive neural networks with different time scales". In: *Neural Computation* 8.8, pp. 1731–1742.

Michalareas, Georgios, Julien Vezoli, Stan Van Pelt, Jan-Mathijs Schoffelen, Henry Kennedy, and Pascal Fries (2016). "Alpha-beta and gamma rhythms subserve feedback and feedforward influences among human visual cortical areas". In: *Neuron* 89.2, pp. 384–397.

Mikolov, Tomas, Martin Karafiát, Lukas Burget, Jan Cernockỳ, and Sanjeev Khudanpur (2010). "Recurrent neural network based language model." In: *Interspeech*. Vol. 2, p. 3.

Milner, Peter M et al. (1974). "A model for visual shape recognition." In: *Psychological review* 81.6, p. 521.

Mnih, Volodymyr, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin Riedmiller (2013). "Playing atari with deep reinforcement learning". In: *arXiv preprint arXiv:1312.5602*.

Moran, Rosalyn J, Stefan J Kiebel, KE Stephan, RB Reilly, Jean Daunizeau, and Karl J Friston (2007). "A neural mass model of spectral responses in electrophysiology". In: *NeuroImage* 37.3, pp. 706–720.

Moreno, Yamir and Amalio F Pacheco (2004). "Synchronization of Kuramoto oscillators in scale-free networks". In: *EPL (Europhysics Letters)* 68.4, p. 603.

Mori, Susumu and Peter van Zijl (2002). "Fiber tracking: principles and strategies–a technical review". In: *NMR in Biomedicine* 15.7-8, pp. 468–480.

Moscovitch, Morris (1995). "Recovered consciousness: A hypothesis concerning modularity and episodic memory". In: *Journal of clinical and experimental neuropsychology* 17.2, pp. 276–290.

Movshon, J Anthony, Ian D Thompson, and David J Tolhurst (1978a). "Spatial summation in the receptive fields of simple cells in the cat's striate cortex." In: *The Journal of physiology* 283, p. 53.

Movshon, J Anthony, ID Thompson, and DJ Tolhurst (1978b). "Receptive field organization of complex cells in the cat's striate cortex." In: *The Journal of physiology* 283, p. 79.

Mozolic, Jennifer L, Christina E Hugenschmidt, Ann M Peiffer, and Paul J Laurienti (2008). "Modality-specific selective attention attenuates multisensory integration". In: *Experimental brain research* 184.1, pp. 39–52.

Murthy, Venkatesh N and Eberhard E Fetz (1996). "Synchronization of neurons during local field potential oscillations in sensorimotor cortex of awake monkeys". In: *Journal of Neurophysiology* 76.6, pp. 3968–3982.

Nadel, Lynn and Morris Moscovitch (1997). "Memory consolidation, retrograde amnesia and the hippocampal complex". In: *Current opinion in neurobiology* 7.2, pp. 217–227.

Nagel, Saskia K, Christine Carl, Tobias Kringe, Robert Märtin, and Peter König (2005). "Beyond sensory substitution—learning the sixth sense". In: *Journal of neural engineering* 2.4, R13.

Nardini, Marko, Peter Jones, Rachael Bedford, and Oliver Braddick (2008). "Development of cue integration in human navigation". In: *Current biology* 18.9, pp. 689–693.

Necker, Louis Albert (1832). "LXI. Observations on some remarkable optical phænomena seen in Switzerland; and on an optical phænomenon which occurs on viewing a figure of a crystal or geometrical solid". In:

Neil, J, J Miller, P Mukherjee, and Petra Susan Hüppi (2002). "Diffusion tensor imaging of normal and injured developing human brain-a technical review". In: *NMR in Biomedicine* 15.7-8, pp. 543–552.

Newman, Mark EJ (2008). "The mathematics of networks". In: *The new palgrave encyclopedia of economics* 2.2008, pp. 1–12.

Ngiam, Jiquan, Adam Coates, Ahbik Lahiri, Bobby Prochnow, Andrew Ng, and Quoc V Le (2011). "On optimization methods for deep learning". In: *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, pp. 265–272.

Niso, Guiomar, Sira Carrasco, María Gudín, Fernando Maestú, Francisco del Pozo, and Ernesto Pereda (2015). "What graph theory actually tells us about resting state interictal MEG epileptic activity". In: *NeuroImage: Clinical*.

Nolte, Guido, Ou Bai, Lewis Wheaton, Zoltan Mari, Sherry Vorbach, and Mark Hallett (2004). "Identifying true brain interaction from EEG data using the imaginary part of coherency". In: *Clinical neurophysiology* 115.10, pp. 2292–2307.

Nunez, Paul L and Ramesh Srinivasan (2006). *Electric fields of the brain: the neurophysics of EEG*. Oxford university press.

Nunez, Paul L, Ramesh Srinivasan, Andrew F Westdorp, Ranjith S Wijesinghe, Don M Tucker, Richard B Silberstein, and Peter J Cadusch (1997). "EEG coherency: I: statistics, reference electrode, volume conduction, Laplacians, cortical imaging, and interpretation at multiple scales". In: *Electroencephalography and clinical neurophysiology* 103.5, pp. 499–515.

Nyhus, Erika and Tim Curran (2010). "Functional role of gamma and theta oscillations in episodic memory". In: *Neuroscience & Biobehavioral Reviews* 34.7, pp. 1023–1035.

Obleser, Jonas, Malte Wöstmann, Nele Hellbernd, Anna Wilsch, and Burkhard Maess (2012). "Adverse listening conditions and memory load drive a common alpha oscillatory network". In: *The Journal of Neuroscience* 32.36, pp. 12376–12383.

O'Keefe, John and Jonathan Dostrovsky (1971). "The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat". In: *Brain research* 34.1, pp. 171–175.

O'keefe, John and Lynn Nadel (1978). *The hippocampus as a cognitive map*. Oxford University Press, USA.

O'Keefe, John and Michael L Recce (1993). "Phase relationship between hippocampal place units and the EEG theta rhythm". In: *Hippocampus* 3.3, pp. 317–330.

Olshausen, Bruno A and David J Field (1997). "Sparse coding with an overcomplete basis set: A strategy employed by V1?" In: *Vision research* 37.23, pp. 3311–3325.

– (2004). "Sparse coding of sensory inputs". In: *Current opinion in neurobiology* 14.4, pp. 481–487.

– (2005). "How close are we to understanding V1?" In: *Neural computation* 17.8, pp. 1665–1699.

Olshausen, Bruno A et al. (1996). "Emergence of simple-cell receptive field properties by learning a sparse code for natural images". In: *Nature* 381.6583, pp. 607–609.

Onat, Selim, Dirk Jancke, and Peter König (2013). "Cortical long-range interactions embed statistical knowledge of natural sensory input: a voltage-sensitive dye imaging study". In: *F1000Research* 2.

Oord, Aaron van den, Nal Kalchbrenner, and Koray Kavukcuoglu (2016). "Pixel Recurrent Neural Networks". In: *arXiv preprint arXiv:1601.06759*.

Oostenveld, Robert, Dick F Stegeman, Peter Praamstra, and Adriaan van Oosterom (2003). "Brain symmetry and topographic analysis of lateralized event-related potentials". In: *Clinical neurophysiology* 114.7, pp. 1194–1202.

Oostenveld, Robert, Pascal Fries, Eric Maris, and Jan-Mathijs Schoffelen (2010). "FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data". In: *Computational intelligence and neuroscience* 2011.

Osipova, Daria, Atsuko Takashima, Robert Oostenveld, Guillén Fernández, Eric Maris, and Ole Jensen (2006). "Theta and gamma oscillations predict encoding and retrieval of declarative memory". In: *The Journal of neuroscience* 26.28, pp. 7523–7531.

Pan, Sinno Jialin and Qiang Yang (2010). "A survey on transfer learning". In: *IEEE Transactions on knowledge and data engineering* 22.10, pp. 1345–1359.

Parise, Cesare V, Charles Spence, and Marc O Ernst (2012). "When correlation implies causation in multisensory integration". In: *Current Biology* 22.1, pp. 46–49.

Pascual-Marqui, Roberto D (2007a). "Discrete, 3D distributed, linear imaging methods of electric neuronal activity. Part 1: exact, zero error localization". In: *arXiv preprint arXiv:0710.3341*.

– (2007b). "Instantaneous and lagged measurements of linear and nonlinear dependence between groups of multivariate time series: frequency decomposition". In: *arXiv preprint arXiv:0711.1455*.

Pascual-Marqui, Roberto D, Dietrich Lehmann, Martha Koukkou, Kieko Kochi, Peter Anderer, Bernd Saletu, Hideaki Tanaka, Koichi Hirata, E Roy John, Leslie Prichep, et al. (2011). "Assessing interactions in the brain with exact low-resolution electromagnetic tomography". In: *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 369.1952, pp. 3768–3784.

Pasquale, Francesco de, Stefania Della Penna, Abraham Z Snyder, Christopher Lewis, Dante Mantini, Laura Marzetti, Paolo Belardinelli, Luca Ciancetta, Vittorio Pizzella, Gian Luca Romani, et al. (2010). "Temporal dynamics of spontaneous MEG activity in brain networks". In: *Proceedings of the National Academy of Sciences* 107.13, pp. 6040–6045.

Paulsen, Ole and Terrence J Sejnowski (2000). "Natural patterns of activity and long-term synaptic plasticity". In: *Current opinion in neurobiology* 10.2, pp. 172–180.

Pelphrey, Kevin A, James P Morris, and Gregory Mccarthy (2004). "Grasping the intentions of others: the perceived intentionality of an action influences activity in the superior temporal sulcus during social perception". In: *Journal of cognitive neuroscience* 16.10, pp. 1706–1716.

Perrett, DI, ET Rolls, and W Caan (1982). "Visual neurones responsive to faces in the monkey temporal cortex". In: *Experimental brain research* 47.3, pp. 329–342.

Peyrache, Adrien, Mehdi Khamassi, Karim Benchenane, Sidney I Wiener, and Francesco P Battaglia (2009). "Replay of rule-learning related neural patterns in the prefrontal cortex during sleep". In: *Nature neuroscience* 12.7, pp. 919–926.

Potter, Mary C, Marvin M Chun, Bradley S Banks, and Margaret Muckenhoupt (1998). "Two attentional deficits in serial target search: the visual attentional blink and an amodal task-switch deficit." In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 24.4, p. 979.

Power, Jonathan D, Alexander L Cohen, Steven M Nelson, Gagan S Wig, Kelly Anne Barnes, Jessica A Church, Alecia C Vogel, Timothy O Laumann, Fran

M Miezin, Bradley L Schlaggar, et al. (2011). "Functional network organization of the human brain". In: *Neuron* 72.4, pp. 665–678.

Raichle, Marcus E, Ann Mary MacLeod, Abraham Z Snyder, William J Powers, Debra A Gusnard, and Gordon L Shulman (2001). "A default mode of brain function". In: *Proceedings of the National Academy of Sciences* 98.2, pp. 676–682.

Raina, Rajat, Alexis Battle, Honglak Lee, Benjamin Packer, and Andrew Y Ng (2007). "Self-taught learning: transfer learning from unlabeled data". In: *Proceedings of the 24th international conference on Machine learning*. ACM, pp. 759–766.

Ranzato, Marc'Aurelio and Geoffrey E Hinton (2010). "Modeling pixel means and covariances using factorized third-order boltzmann machines". In: *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, pp. 2551–2558.

Rappelsberger, P, G Pfurtscheller, and O Filz (1994). "Calculation of event-related coherence—a new method to study short-lasting coupling between brain areas". In: *Brain topography* 7.2, pp. 121–127.

Rasch, Björn and Jan Born (2013). "About sleep's role in memory". In: *Physiological reviews* 93.2, pp. 681–766.

Rasch, Björn, Christian Büchel, Steffen Gais, and Jan Born (2007). "Odor cues during slow-wave sleep prompt declarative memory consolidation". In: *Science* 315.5817, pp. 1426–1429.

Rauschecker, Joseph P and Wolf Singer (1981). "The effects of early visual experience on the cat's visual cortex and their possible explanation by Hebb synapses." In: *The Journal of physiology* 310, p. 215.

Rehn, Martin and Friedrich T Sommer (2007). "A network that uses few active neurones to code visual input predicts the diverse shapes of cortical receptive fields". In: *Journal of computational neuroscience* 22.2, pp. 135–146.

Reichert, David P and Thomas Serre (2013). "Neuronal synchrony in complex-valued deep networks". In: *arXiv preprint arXiv:1312.6115*.

Remondes, Miguel and Erin M Schuman (2004). "Role for a cortical input to hippocampal area CA1 in the consolidation of a long-term memory". In: *Nature* 431.7009, pp. 699–703.

Reuschel, Johanna, Knut Drewing, Denise YP Henriques, Frank Rösler, and Katja Fiehler (2010). "Optimal integration of visual and proprioceptive movement information for the perception of trajectory geometry". In: *Experimental brain research* 201.4, pp. 853–862.

Reuter, Martin, Nicholas J Schmansky, H Diana Rosas, and Bruce Fischl (2012). "Within-subject template estimation for unbiased longitudinal image analysis". In: *Neuroimage* 61.4, pp. 1402–1418.

Ritter, Petra, Michael Schirner, Anthony R McIntosh, and Viktor K Jirsa (2013). "The virtual brain integrates computational modeling and multimodal neuroimaging". In: *Brain connectivity* 3.2, pp. 121–145.

Roelfsema, Pieter R, Andreas K Engel, Peter König, and Wolf Singer (1997). "Visuomotor integration is associated with zero time-lag synchronization among cortical areas". In: *Nature* 385, pp. 157–161.

Rubinov, Mikail and Olaf Sporns (2010). "Complex network measures of brain connectivity: uses and interpretations". In: *Neuroimage* 52.3, pp. 1059–1069.

Ruder, Sebastian (2016). "An overview of gradient descent optimization algorithms". In: *arXiv preprint arXiv:1609.04747*.

Rumbell, Timothy, Susan L Denham, and Thomas Wennekers (2014). "A spiking self-organizing map combining stdp, oscillations, and continuous learning". In: *IEEE transactions on neural networks and learning systems* 25.5, pp. 894–907.

Russell, Bryan C, Antonio Torralba, Kevin P Murphy, and William T Freeman (2008). "LabelMe: a database and web-based tool for image annotation". In: *International journal of computer vision* 77.1-3, pp. 157–173.

Sadilek, Maximilian and Stefan Thurner (2015). "Physiologically motivated multiplex Kuramoto model describes phase diagram of cortical activity". In: *Scientific reports* 5.

Samsonovich, Alexei and Bruce L McNaughton (1997). "Path integration and cognitive mapping in a continuous attractor neural network model". In: *The Journal of neuroscience* 17.15, pp. 5900–5920.

Schirner, Michael, Simon Rothmeier, Viktor K Jirsa, Anthony Randal McIntosh, and Petra Ritter (2015). "An automated pipeline for constructing personalized virtual brains from multimodal neuroimaging data". In: *NeuroImage*.

Schmahmann, Jeremy D and Deepak Pandya (2009). *Fiber pathways of the brain*. OUP USA.

Schmidhuber, Jürgen (2015). "Deep learning in neural networks: An overview". In: *Neural Networks* 61, pp. 85–117.

Schoffelen, Jan-Mathijs and Joachim Gross (2009). "Source connectivity analysis with MEG and EEG". In: *Human brain mapping* 30.6, pp. 1857–1865.

Scoville, William Beecher and Brenda Milner (1957). "Loss of recent memory after bilateral hippocampal lesions". In: *Journal of Neurology, Neurosurgery & Psychiatry* 20.1, pp. 11–21.

Sederberg, Per B, Michael J Kahana, Marc W Howard, Elizabeth J Donner, and Joseph R Madsen (2003). "Theta and gamma oscillations during encoding predict subsequent recall". In: *The Journal of Neuroscience* 23.34, pp. 10809–10814.

Shams, Ladan and Aaron R Seitz (2008). "Benefits of multisensory learning". In: *Trends in cognitive sciences* 12.11, pp. 411–417.

Shi, Jianbo and Jitendra Malik (2000). "Normalized cuts and image segmentation". In: *IEEE Transactions on pattern analysis and machine intelligence* 22.8, pp. 888–905.

Siegel, Markus, Tobias H Donner, and Andreas K Engel (2012). "Spectral fingerprints of large-scale neuronal interactions". In: *Nature Reviews Neuroscience* 13.2, pp. 121–134.

Silva, Laurie R, Yael Amitai, and Barry W Connors (1991). "Intrinsic oscillations of neocortex generated by layer 5 pyramidal neurons". In: *Science* 251.4992, p. 432.

Simmonds, Daniel J, James J Pekar, and Stewart H Mostofsky (2008). "Meta-analysis of Go/No-go tasks demonstrating that fMRI activation associated with response inhibition is task-dependent". In: *Neuropsychologia* 46.1, pp. 224–232.

Simoncelli, Eero P and Bruno A Olshausen (2001). "Natural image statistics and neural representation". In: *Annual review of neuroscience* 24.1, pp. 1193–1216.

Simonyan, Karen and Andrew Zisserman (2014). "Very deep convolutional networks for large-scale image recognition". In: *arXiv preprint arXiv:1409.1556*.

Singer, Wolf (1999). "Neuronal synchrony: a versatile code for the definition of relations?" In: *Neuron* 24.1, pp. 49–65.

– (2004). "Synchrony, oscillations, and relational codes". In: *The visual neurosciences* 2, pp. 1665–1681.

Singer, Wolf and Charles M Gray (1995). "Visual feature integration and the temporal correlation hypothesis". In: *Annual review of neuroscience* 18.1, pp. 555–586.

Sirota, Anton, Sean Montgomery, Shigeyoshi Fujisawa, Yoshikazu Isomura, Michael Zugaro, and György Buzsáki (2008). "Entrainment of neocortical

neurons and gamma oscillations by the hippocampal theta rhythm". In: *Neuron* 60.4, pp. 683–697.

Skaggs, William E and Bruce L McNaughton (1996a). "Replay of neuronal firing sequences in rat hippocampus during sleep following spatial experience". In: *Science* 271.5257, p. 1870.

– (1996b). "Theta phase precession in hippocampal". In: *Hippocampus* 6, pp. 149–172.

Skudlarski, Pawel, Kanchana Jagannathan, Vince D Calhoun, Michelle Hampson, Beata A Skudlarska, and Godfrey Pearlson (2008). "Measuring brain connectivity: diffusion tensor imaging validates resting state temporal correlations". In: *Neuroimage* 43.3, pp. 554–561.

Smith, Barry (1988). "Foundations of Gestalt theory". In:

Snodgrass, Joan G and June Corwin (1988). "Pragmatics of measuring recognition memory: applications to dementia and amnesia." In: *Journal of Experimental Psychology: General* 117.1, p. 34.

Sompolinsky, H, D Golomb, and D Kleinfeld (1990). "Global processing of visual stimuli in a neural network of coupled oscillators". In: *Proceedings of the National Academy of Sciences* 87.18, pp. 7200–7204.

Song, Sen, Kenneth D Miller, and Larry F Abbott (2000). "Competitive Hebbian learning through spike-timing-dependent synaptic plasticity". In: *Nature neuroscience* 3.9, pp. 919–926.

Spall, James C (2005). *Introduction to stochastic search and optimization: estimation, simulation, and control*. Vol. 65. John Wiley & Sons.

Sporns, Olaf, Giulio Tononi, and Gerald M Edelman (2000). "Theoretical neuroanatomy: relating anatomical and functional connectivity in graphs and cortical connection matrices". In: *Cerebral Cortex* 10.2, pp. 127–141.

Sporns, Olaf, Giulio Tononi, and Rolf Kötter (2005). "The human connectome: a structural description of the human brain". In: *PLoS Comput Biol* 1.4, e42.

Sporns, Olaf, Dante R Chialvo, Marcus Kaiser, Claus C Hilgetag, et al. (2004). "Organization, development and function of complex brain networks". In: *Trends in cognitive sciences* 8.9, pp. 418–425.

Squire, Larry R and John T Wixted (2011). "The cognitive neuroscience of human memory since HM". In: *Annual review of neuroscience* 34, p. 259.

Squire, Larry R and Stuart M Zola (1997). "Amnesia, memory and brain systems". In: *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 352.1362, pp. 1663–1673.

Srivastava, Nitish, Geoffrey E Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov (2014). "Dropout: a simple way to prevent neural networks from overfitting." In: *Journal of Machine Learning Research* 15.1, pp. 1929–1958.

Stam, Cornelis J, Guido Nolte, and Andreas Daffertshofer (2007). "Phase lag index: Assessment of functional connectivity from multi channel EEG and MEG with diminished bias from common sources". In: *Human brain mapping* 28.11, pp. 1178–1193.

Staudigl, Tobias and Simon Hanslmayr (2013). "Theta oscillations at encoding mediate the context-dependent nature of human episodic memory". In: *Current Biology* 23.12, pp. 1101–1106.

Stein, Barry E and Terrence R Stanford (2008). "Multisensory integration: current issues from the perspective of the single neuron". In: *Nature Reviews Neuroscience* 9.4, pp. 255–266.

Steinmann, Saskia, Gregor Leicht, Matthias Ertl, Christina Andreou, Nenad Polomac, René Westerhausen, Angela D Friederici, and Christoph Mulert (2014). "Conscious auditory perception related to long-range synchrony of gamma oscillations". In: *NeuroImage* 100, pp. 435–443.

Steriade, Mircea, A Nunez, and F Amzica (1993). "A novel slow (<1 Hz) oscillation of neocortical neurons in vivo: depolarizing and hyperpolarizing components". In: *The Journal of Neuroscience* 13.8, pp. 3252–3265.

Stickgold, Robert (2005). "Sleep-dependent memory consolidation". In: *Nature* 437.7063, pp. 1272–1278.

Stimberg, Marcel, Klaus Wimmer, Robert Martin, Lars Schwabe, Jorge Mariño, James Schummers, David C Lyon, Mriganka Sur, and Klaus Obermayer (2009). "The operating regime of local computations in primary visual cortex". In: *Cerebral cortex* 19.9, pp. 2166–2180.

Stinstra, JG and MJ Peters (1998). "The volume conductor may act as a temporal filter on the ECG and EEG". In: *Medical and Biological Engineering and Computing* 36.6, pp. 711–716.

Stryker, Michael P, Helen Sherk, Audie G Leventhal, and Helmut V Hirsch (1978). "Physiological consequences for the cat's visual cortex of effectively restricting early visual experience with oriented contours". In: *Journal of Neurophysiology* 41.4, pp. 896–909.

Sturm, Anja K and Peter König (2001). "Mechanisms to synchronize neuronal activity". In: *Biological cybernetics* 84.3, pp. 153–172.

Supekar, Kaustubh, Lucina Q Uddin, Katherine Prater, Hitha Amin, Michael D Greicius, and Vinod Menon (2010). "Development of functional and structural connectivity within the default mode network in young children". In: *Neuroimage* 52.1, pp. 290–301.

Sutherland, Gary R and Bruce McNaughton (2000). "Memory trace reactivation in hippocampal and neocortical neuronal ensembles". In: *Current opinion in neurobiology* 10.2, pp. 180–186.

Sutskever, Ilya, James Martens, George E Dahl, and Geoffrey E Hinton (2013). "On the importance of initialization and momentum in deep learning." In: *ICML (3)* 28, pp. 1139–1147.

Swindale, NV (1982). "A model for the formation of orientation columns". In: *Proceedings of the Royal Society of London B: Biological Sciences* 215.1199, pp. 211–230.

Szegedy, Christian, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich (2015). "Going deeper with convolutions". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–9.

Takashima, A, Karl Magnus Petersson, F Rutters, I Tendolkar, O Jensen, MJ Zwarts, BL McNaughton, and G Fernandez (2006). "Declarative memory consolidation in humans: a prospective functional magnetic resonance imaging study". In: *Proceedings of the National Academy of Sciences of the United States of America* 103.3, pp. 756–761.

Tallon-Baudry, Catherine and Olivier Bertrand (1999). "Oscillatory gamma activity in humans and its role in object representation". In: *Trends in cognitive sciences* 3.4, pp. 151–162.

Talsma, Durk, Tracy J Doty, and Marty G Woldorff (2007). "Selective attention and audiovisual integration: is attending to both modalities a prerequisite for early integration?" In: *Cerebral cortex* 17.3, pp. 679–690.

Tetko, Igor V, David J Livingstone, and Alexander I Luik (1995). "Neural network studies. 1. Comparison of overfitting and overtraining". In: *Journal of chemical information and computer sciences* 35.5, pp. 826–833.

Tieleman, Tijmen and Geoffrey Hinton (2012). "Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude". In: *COURSERA: Neural Networks for Machine Learning* 4.2.

Ton, Robert, Gustavo Deco, and Andreas Daffertshofer (2014). "Structure-Function Discrepancy: Inhomogeneity and Delays in Synchronized Neural Networks." In: *PLoS Computational Biology* 10.7.

Tononi, Giulio and Chiara Cirelli (2006). "Sleep function and synaptic homeostasis". In: *Sleep medicine reviews* 10.1, pp. 49–62.

Tononi, Giulio, Olaf Sporns, and Gerald M Edelman (1994). "A measure for brain complexity: relating functional segregation and integration in the nervous system". In: *Proceedings of the National Academy of Sciences* 91.11, pp. 5033–5037.

Tononi, Giullo, Olaf Sporns, and Gerald M Edelman (1992). "Reentry and the problem of integrating multiple cortical areas: simulation of dynamic integration in the visual system". In: *Cerebral Cortex* 2.4, pp. 310–335.

Tournier, J, Fernando Calamante, Alan Connelly, et al. (2012). "MRtrix: diffusion tractography in crossing fiber regions". In: *International Journal of Imaging Systems and Technology* 22.1, pp. 53–66.

Treves, Alessandro and Edmund T Rolls (1994). "Computational analysis of the role of the hippocampus in memory". In: *Hippocampus* 4.3, pp. 374–391.

Tsukada, Koji and Michiaki Yasumura (2004). "Activebelt: Belt-type wearable tactile display for directional navigation". In: *International Conference on Ubiquitous Computing*. Springer, pp. 384–399.

Van Erp, Jan BF and HAHC Van Veen (2003). "A multi-purpose tactile vest for astronauts in the international space station". In: *Proceedings of eurohaptics*. Dublin, Ireland: ACM, Press, pp. 405–408.

Van Veen, Barry D, Wim Van Drongelen, Moshe Yuchtman, and Akifumi Suzuki (1997). "Localization of brain electrical activity via linearly constrained minimum variance spatial filtering". In: *Biomedical Engineering, IEEE Transactions on* 44.9, pp. 867–880.

Váša, František, Murray Shanahan, Peter J Hellyer, Gregory Scott, Joana Cabral, and Robert Leech (2015). "Effects of lesions on synchrony and metastability in cortical networks". In: *NeuroImage*.

Vecchio, Fabrizio, Francesca Miraglia, Giuseppe Curcio, Giacomo Della Marca, Catello Vollono, Edoardo Mazzucchi, Placido Bramanti, and Paolo Maria Rossini (2015). "Cortical connectivity in fronto-temporal focal epilepsy from EEG analysis: A study via graph theory". In: *Clinical Neurophysiology* 126.6, pp. 1108–1116.

Victor, Jonathan D and Keith P Purpura (1996). "Nature and precision of temporal coding in visual cortex: a metric-space analysis". In: *Journal of neurophysiology* 76.2, pp. 1310–1326.

Vilalta, Ricardo and Youssef Drissi (2002). "A perspective view and survey of meta-learning". In: *Artificial Intelligence Review* 18.2, pp. 77–95.

Villegas, Pablo, Paolo Moretti, and Miguel A Muñoz (2014). "Frustrated hierarchical synchronization and emergent complexity in the human connectome network". In: *Scientific reports* 4.

Vincent, JL, GH Patel, MD Fox, AZ Snyder, JT Baker, DC Van Essen, JM Zempel, LH Snyder, M Corbetta, and ME Raichle (2007). "Intrinsic functional architecture in the anaesthetized monkey brain". In: *Nature* 447.7140, pp. 83–86.

Vincent, Pascal, Hugo Larochelle, Yoshua Bengio, and Pierre-Antoine Manzagol (2008). "Extracting and composing robust features with denoising autoencoders". In: *Proceedings of the 25th international conference on Machine learning*. ACM, pp. 1096–1103.

Vincent, Pascal, Hugo Larochelle, Isabelle Lajoie, Yoshua Bengio, and Pierre-Antoine Manzagol (2010). "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion". In: *Journal of Machine Learning Research* 11.Dec, pp. 3371–3408.

Vinck, Martin, Robert Oostenveld, Marijn van Wingerden, Franscesco Battaglia, and Cyriel MA Pennartz (2011). "An improved index of phase-synchronization for electrophysiological data in the presence of volume-conduction, noise and sample-size bias". In: *Neuroimage* 55.4, pp. 1548–1565.

Vindiola, MM, JM Vettel, SM Gordon, PJ Franaszczuk, and K McDowell (2014). "Applying EEG phase synchronization measures to non-linearly coupled neural mass models". In: *Journal of neuroscience methods* 226, pp. 1–14.

Vyazovskiy, Vladyslav V, Chiara Cirelli, Martha Pfister-Genskow, Ugo Faraguna, and Giulio Tononi (2008). "Molecular and electrophysiological evidence for net synaptic potentiation in wake and depression in sleep". In: *Nature neuroscience* 11.2, pp. 200–208.

Wahn, Basil and Peter König (2015a). "Audition and vision share spatial attentional resources, yet attentional load does not disrupt audiovisual integration". In: *Frontiers in psychology* 6.

– (2015b). "Vision and haptics share spatial attentional resources and visuotactile integration is not affected by high attentional load". In: *Multisensory research* 28.3-4, pp. 371–392.

– (2016). “Attentional resource allocation in visuotactile processing depends on the task, but optimal visuotactile integration does not depend on attentional resources”. In: *Frontiers in integrative neuroscience* 10.

Wang, DeLiang and David Terman (1997). “Image segmentation based on oscillatory correlation”. In: *Neural Computation* 9.4, pp. 805–836.

Watts, Duncan J and Steven H Strogatz (1998). “Collective dynamics of ‘small-world’ networks”. In: *nature* 393.6684, pp. 440–442.

Wedeen, Van J, RP Wang, Jeremy D Schmahmann, T Benner, WYI Tseng, Guangping Dai, DN Pandya, Patric Hagmann, Helen D’Arceuil, and Alex J de Crespigny (2008). “Diffusion spectrum magnetic resonance imaging (DSI) tractography of crossing fibers”. In: *Neuroimage* 41.4, pp. 1267–1277.

Wiggs, Cheri L and Alex Martin (1998). “Properties and mechanisms of perceptual priming”. In: *Current opinion in neurobiology* 8.2, pp. 227–233.

Wilson, Hugh R and Jack D Cowan (1972). “Excitatory and inhibitory interactions in localized populations of model neurons”. In: *Biophysical journal* 12.1, p. 1.

Wilson, Matthew A, Bruce L McNaughton, et al. (1994). “Reactivation of hippocampal ensemble memories during sleep”. In: *Science* 265.5172, pp. 676–679.

Wise, Roy A (2004). “Dopamine, learning and motivation”. In: *Nature reviews neuroscience* 5.6, pp. 483–494.

Wiskott, Laurenz and Terrence J Sejnowski (2002). “Slow feature analysis: Unsupervised learning of invariances”. In: *Neural computation* 14.4, pp. 715–770.

Wolf, Fred and Theo Geisel (1998). “Spontaneous pinwheel annihilation during visual development”. In: *Nature* 395.6697, pp. 73–78.

Womelsdorf, Thilo, Jan-Mathijs Schoffelen, Robert Oostenveld, Wolf Singer, Robert Desimone, Andreas K Engel, and Pascal Fries (2007). “Modulation of neuronal interactions through neuronal synchronization”. In: *science* 316.5831, pp. 1609–1612.

Woolrich, Mark W and Klaas E Stephan (2013). “Biophysical network models and the human connectome”. In: *Neuroimage* 80, pp. 330–338.

Wozny, David R, Ulrik R Beierholm, and Ladan Shams (2010). “Probability matching as a computational strategy used in perception”. In: *PLoS Comput Biol* 6.8, e1000871.

Wu, Jennifer, Ramesh Srinivasan, Arshdeep Kaur, and Steven C Cramer (2014). "Resting-state cortical connectivity predicts motor skill acquisition". In: *NeuroImage* 91, pp. 84–90.

Wyss, Reto, Peter König, and Paul FM J Verschure (2006). "A model of the ventral visual system based on temporal stability and local memory". In: *PLoS biology* 4.5, e120.

Yeung, MK Stephen and Steven H Strogatz (1999). "Time delay in the Kuramoto model of coupled oscillators". In: *Physical Review Letters* 82.3, p. 648.

Young, Richard A (1987). "The Gaussian derivative model for spatial vision: I. Retinal mechanisms". In: *Spatial Vision* 2.4, pp. 273–293.

Young, Richard A and Ronald M Lesperance (2001). "The gaussian derivative model for spatial-temporal vision: II. Cortical data". In: *Spatial vision* 14.3-4, pp. 3–4.

Yuval-Greenberg, Shlomit, Orr Tomer, Alon S Keren, Israel Nelken, and Leon Y Deouell (2008). "Transient induced gamma-band response in EEG as a manifestation of miniature saccades". In: *Neuron* 58.3, pp. 429–441.

Zalesky, Andrew, Alex Fornito, Ian H Harding, Luca Cocchi, Murat Yücel, Christos Pantelis, and Edward T Bullmore (2010). "Whole-brain anatomical networks: does the choice of nodes matter?" In: *Neuroimage* 50.3, pp. 970–983.

Zhang, Chong, Xiaolin Yu, Yong Yang, and Lei Xu (2014). "Phase Synchronization and Spectral Coherence Analysis of EEG Activity During Mental Fatigue". In: *Clinical EEG and neuroscience*, p. 1550059413503961.

Zhang, Guoqiang Peter (2000). "Neural networks for classification: a survey". In: *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 30.4, pp. 451–462.

Zhong, Suyu, Yong He, and Gaolang Gong (2015). "Convergence and divergence across construction methods for human brain white matter networks: An assessment based on individual differences". In: *Human brain mapping* 36.5, pp. 1995–2013.

Adaptability and flexibility are some of the most important human characteristics. Learning based on new experiences enables adaptation by changing the structural connectivity of the brain through plasticity mechanisms. But the human brain can also adapt to new tasks and situations in a matter of milliseconds by dynamic coordination of functional activation. To understand how this flexibility can be achieved in the computations performed by neural networks, we have to understand how the relatively fixed structural backbone interacts with the functional dynamics. In this thesis, I will analyze these interactions between the structural network connectivity and functional activations and their dynamic interactions on different levels of abstraction and spatial and temporal scales.

Four studies are presented in this work:

(1) A connectome modeling study investigating several alternative methods along the modeling path from the estimation of structural connectivity via computational models to the estimation of functional connectivity from EEG recordings.

(2) A temporal coding study that extends the rate-based units in models of representation learning from natural visual stimuli with phase oscillators to investigate binding by synchrony.

(3) A sensory integration study investigating integration of rotation cues of an augmented tactile modality and the native vestibular modality.

(4) A memory consolidation study investigating the effects of sleep on the cortical oscillatory interactions during the retrieval of associative memory items.