



HAL
open science

The Role of neuronal oscillations in local computations and network brain dynamics

Tommy Clausner

► **To cite this version:**

Tommy Clausner. The Role of neuronal oscillations in local computations and network brain dynamics. Neuroscience. Université Claude Bernard Lyon 1, 2022. English. NNT: 2022LYO1242 . tel-04040022

HAL Id: tel-04040022

<https://hal.science/tel-04040022>

Submitted on 21 Mar 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



N°d'ordre NNT :

2022LYO1242

THESE de DOCTORAT DE L'UNIVERSITE DE LYON

opérée au sein de

l'Université Claude Bernard Lyon 1

**Ecole Doctorale N° 476 - NSCO
ECOLE DOCTORALE NEUROSCIENCE ET COGNITION**

Spécialité de doctorat :

Neurosciences

Soutenue publiquement le 15/12/2022, par :

Tommy Clausner

**The Role of neuronal oscillations in local
computations and network brain dynamics**

Devant le jury composé de :

Président.e (à préciser après la soutenance)

Hanslmayr, Simon Professor University of Glasgow
de Lange, Floris Professor Radboud University Nijmegen
Zumer, Johanna Lecturer Aston University, Birmingham
Scheeringa, Réne Researcher Radboud University Nijmegen
Gervais, Rémi Professor emeritus University Lyon 1
Bonfond, Mathilde principle investigator Inserm Lyon

Rapporteur
Rapporteur
Examinatrice
Examineur
Examineur
Directrice de thèse

The Role of neuronal oscillations in local computations and network brain dynamics

Le rôle des oscillations neuronales dans les processus computationnels locaux et la dynamique des réseaux cérébraux

by Tommy CLAUSNER

These de doctorat de l'Université de Lyon
PhD thesis

Université Claude Bernard  Lyon 1

Ecole Doctorale N° 476 - NSCO
Ecole Doctorale Neurosciences et Cognition
l'Université Claude Bernard Lyon 1
France

thesis director | Mathilde BONNEFOND (INSERM, Lyon)

reviewer | Simon HANSLMAYR (University of Glasgow)
reviewer | Floris DE LANGE (Donders Institute, Nijmegen)
examiner | Johanna ZUMER (Aston University, Birmingham)
examiner | René SCHEERINGA (Donders Institute, Nijmegen)
examiner | Rémi GERVAIS, (University Lyon 1)

January 15, 2023

N°d'ordre NNT: 2022LYO1242

Université Claude Bernard – Lyon 1

Président de l'Université	Frédéric FLEURY
Président du Conseil Académique	Hamda BEN HADID
Vice-Président du Conseil d'Administration	Didier REVEL
Vice-Président du Conseil des Etudes et de la Vie Universitaire	Céline BROCHIER
Vice-Président de la Commission de Recherche	Petru MIRONESCU
Directeur Général des Services	Pierre ROLLAND

Composantes Sante

Département de Formation et Centre de Recherche en Biologie Humaine	Directrice: Anne-Marie SCHOTT
Faculté d'Odontologie	Doyenne: Dominique SEUX
Faculté de Médecine et Maïeutique Lyon Sud - Charles Mérieux	Doyenne: Carole BURILLON
Faculté de Médecine Lyon-Est	Doyen: Gilles RODE
Institut des Sciences et Techniques de la Réadaptation (ISTR)	Directeur: Xavier PERROT
Institut des Sciences Pharmaceutiques et Biologiques (ISBP)	Directeur: Claude DUSSART

Composantes & Departments de Science & Technologie

Département Génie Electrique et des Procédés (GEP)	Directrice : Mme Rosaria FERRIGNO
Département Informatique	Directeur : M. Behzad SHARIAT
Département Mécanique	Directeur M. Marc BUFFAT
Ecole Supérieure de Chimie, Physique, Electronique (CPE Lyon)	Directeur : Gérard PIGNAULT
Institut de Science Financière et d'Assurances (ISFA)	Directeur : M. Nicolas LEBOISNE
Institut National du Professorat et de l'Éducation	Directeur : M. Pierre CHAREYRON
Institut Universitaire de Technologie de Lyon 1	Directeur : M. Christophe VITON
Observatoire de Lyon	Directrice : Mme Isabelle DANIEL
Polytechnique Lyon	Directeur : Emmanuel PERRIN
UFR Biosciences	Administratrice provisoire : Mme Kathrin GIESELER
UFR des Sciences et Techniques des Activités Physiques et Sportives (STAPS)	Directeur : M. Yannick VANPOULLE
UFR Faculté des Sciences	Directeur : M. Bruno ANDRIOLETTI

General information, publications & financial support

This research was conducted within the Computation, Cognition and Neurophysiology (CO-PHY) team of the Lyon Neuroscience Research Center (CRNL; INSERM U 1028, CNRS UMR 5292, Université Claude Bernard Lyon 1), FR. It was further performed partly in collaboration with the Donders Institute for Brain, Cognition and Behaviour (DCCN), Radboud University, Nijmegen, NL.

This research was funded by the European Research Council under the European Union's Seventh Framework Programme (FP7/2007–2013) / ERC starting grant agreement no 716862 attributed to my PhD supervisor Mathilde Bonnefond. I received financial support from the Fondation pour la Recherche Médicale - grant ID FDT202106013010 in form of a "Fin de These" grant, awarded for the period between October 1, 2021 and September 30, 2022.

This thesis led to the following **publications**:

Clausner, T., & Gentili, S. (2022). Auto-regressive Rank Order Similarity (aros) test. (submitted, preprint on bioRxiv).

Bonnefond M., Jenson O., & Clausner T. Visual processing by dynamic phase coding. (submitted)

Clausner, T., van Mourik, T., Marques, J., Haak, K., & Scheeringa, R.*, Bonnefond, M.* Feature specific neuronal oscillations in the α and γ band are differentially linked to cortical layers. (in prep)

Ferez M., Clausner T., Lukacs J., Gbadoe M., Daligault S., Schwartz D., & Bonnefond M. Evidence of functional inhibition of high visual regions by alpha oscillations (in prep)

and **planned publications** sought as part of a post-doctoral research project:

One first author publication on laminar MEG with respect to attention modulation and predictability (Experiment 1, see page 123); one first author publication on laminar MEG with respect to higher / lower order visual region communication (Experiment 2, see page 130); one second author publication on MEG stimulus decoding with respect to attention and predictability with Maryam Mostafalu [1] as first author (see page 123).

One first author publication on spatial features, learned from γ band power changes by a convolutional neural network (see page 154); one first author publication on oscillatory activity and related "behavioral" changes in a trained spiking neural network (see page 155).

Dedication

This thesis is dedicated to my grandmother Brunhilde Clausner, who has shown unconditional support for me and my work throughout my entire life. Without her support, writing down those words would have been metaphorically - but even literally - much more difficult. In the following paragraph I would hence like to address her personally in German:



Liebe Oma Bruni,

Ich möchte meine Doktorarbeit Dir widmen, denn ich bin sicher, ohne Dich würde es diesen Text vermutlich nicht geben. Mein ganzes Leben lang warst Du immer für mich da und hast mich unterstützt wo Du nur konntest. Dafür werde ich Dir für immer dankbar sein.

Dein Tommy

I would like to thank all members of the committee for accepting to examine my thesis. Many thanks to Simon Hanslmayr and Floris de Lange for reviewing the manuscript and being part of the jury. Additionally I would like to thank Johanna Zumer and Rémi Gervais for agreeing to be part of the examination committee as well. Your expert opinion will not only help me to reflect upon past but furthermore will help me to improve future work, for which I am very grateful!

The final jury member is the supervisor of my doctoral thesis, Mathilde Bonnefond. It is a great honour and pleasure for me to be supervised by such an outstanding person. Not only her exceptionally friendly, humorous and open personality, but also her kind and understanding way of supervision has left a very strong impression on me. I don't know if there was a single discussion with her where I didn't learn at least a thing. More importantly, her academic opinion is of utmost importance and value to me. The combination of rigorous theoretical deduction, the courage to leave well-worn paths, and the creativity needed to push scientific boundaries in meaningful ways are qualities I deeply admire and aspire to emulate. She has certainly become a scientific *and* personal role model to whom I bow my head.

Thank you so much Mathilde!

Moreover, this work stands on the shoulders of many giants, be they friends, family or colleagues. First, I would like to thank René Scheeringa very much for the many hours of very fruitful discussions and support. You have been one of the main contributors to the present project and your expertise has been of immeasurable value. Not to forget that it was you, René, who sparked my interest in the specific topic in the first place! Thank you!

In addition to that, many thanks to the people of our lab, Fardin, Rasa, Maxime, Anne, Julia and Melinda for the so many fruitful discussions, funny moments and great support! Many thanks as well to the awesome interns Lucie and Lea who not only supported me with data collection and analyses but furthermore their curious and open scientific mind made it a bliss having them as short term team members. Furthermore, I want to thank all the other people who helped me to develop my scientific thinking further. Thank you so much Jérémie, Olivier, Nicolas, Christian for your very fruitful feedback during the annual CST meeting. In addition to that, I would like to thank Denis, Sébastien and Franck for their great support during data collection, their patience, helpful attitude and welcoming personalities. In addition to that I'd like to thank PierreM, Thibaut and Hervé for their support with all IT infrastructure related issues. Furthermore, merci Martine, d'être une si bonne secrétaire ! Merci de rendre les choses possibles ! Also a big thanks to Gaëlle who recently took over from Martine and who since then has helped me already a lot with all this messed up French bureaucracy.

On extension I want to thank all those incredibly inspiring people I was allowed to meet, receive their scientific advice, and enjoy their great kindness and support. I would like to thank Wolfgang Klimesch (University of Salzburg); Maite Crespo Garcia, Sarang Dalal, Britta Westner, Monika Zeiller, Gerd Waldhauser, Marie Fellner, Tobias Staudigl, Christian Wienbruch, Brigitte Rockstroh, Dagmar Moret (University of Konstanz); Robert Oostenveld, Jan Mathijs Schoffelen, Paul Gaalman, Hong Lee, Marek Tyc, Mike van Engelenburg, Matthias Ekman (DCCN Nijmegen); Christopher Bailey (Aarhus University).

On a very personal level I want to thank my younger sister Francy (Hallo Francy!) who has been an anchor throughout my entire adult life. Furthermore, thank you Wibke, Thomas, Carl, Stefi for all the many years that you stood so close by my side and also thank you Fia, Robert, Steven, Patricia for all your warmth and inspiration. I wouldn't be the same without you. Furthermore, thanks to all those many people that shared parts of their life with me at one of my stopovers.

Finally, I want to thank my partner Johanna, who has accompanied this journey through every single up and down. There is no way how I could express how deeply grateful I am for her unconditional support. You have been my translator, psychotherapist, life coach, muse, companion and friend. All of this in full time and all of this on the side of being a great partner. With your endless optimism and your boundless kindness you have always been a source of strength and inspiration. Not even a neuro-scientist could measure how much every day with you means to me and wherever we end up in the future, with you it will be

home.

I am incredibly grateful to receive such an outstanding opportunity and I am well aware that without the great support of all those amazing people, this would not have been possible.

Thank you!



English summary

The human brain is a vastly expensive information processing machine with respect to its relative energy uptake compared to other organs. This cost must be outweighed by potential benefits in order to be a viable solution. Hence, it can be hypothesized that the human brain is an extremely efficient information processing apparatus that outperforms any artificial system in terms of energy efficiency by far. Thereby, an understanding of general underlying computational principles not only reveals potential new insights into how artificial neural network systems can be improved but furthermore would have a major impact on clinical applications, such as brain computer interfaces (BCI), artificial sensory organs (e.g. cochlea implants) or robotic prosthetics that communicate directly with the nervous system. It has been demonstrated countless times that neuronal oscillations - rhythmically synchronized neuronal activity - play an important role in cortical signal processing. However, their exact role with respect to the actual ongoing information processing remains still widely unclear. A major problem thereby is that neuronal oscillations and respective cortical computation models make predictions on the level of cortical laminae. While relatively easy accessible in animal models, cortical activity with laminar level resolution ($< 1\text{ mm}$) is difficult to obtain in healthy participants. The present thesis not only aims to demonstrate how to bridge the gap between human and animal model methodologically by implementing state of the art neuro-imaging pipelines, but furthermore targets core predictions derived from animal models and patient studies with respect to the role of neuronal oscillations in local computations and network brain dynamics. A simultaneous EEG-fMRI experiment was conducted with a voxel size of 0.8 mm . Thereby, stimulus feature specific α activity could be shown to differentially respond to preferred compared to not preferred stimuli predominantly in deep and middle cortical layers of V1, separately for low and high α . γ band oscillations have been found to be correlated with the feature processing itself, predominantly in deep and superficial layers. Both findings are in line with previous literature. Additionally two laminar level MEG experiments have been conducted in order to refine respective findings from the previous experiment. Thereby, one MEG experiment has been specifically designed to target low level cortical interactions, whereas the second has been designed to target low to higher order communication for multiple frequency bands. Due to the pandemic and other circumstances, the recordings of the MEG data have been delayed by more than a year. For this reason only preliminary results can be presented. Additionally, it has been planned to investigate different possibilities to test respective hypotheses using neural network simulations. However, as expected, only the initial planning and piloting phase has been reached. In summary, this thesis provides new insights into differential α and γ band activity in different cortical layers, confirming predictions from animal and theoretical models.

Résumé français

Le cerveau humain est une machine de traitement de l'information extrêmement coûteuse en termes de consommation d'énergie par rapport à d'autres organes. Pour être une solution viable, ce coût doit être compensé par les avantages potentiels. On peut donc émettre l'hypothèse que le cerveau humain est un appareil de traitement de l'information extrêmement efficace qui surpasse de loin tout système artificiel en terme d'efficacité énergétique. Ainsi, la compréhension des principes généraux de calcul sous-jacents ne révèle pas seulement de nouvelles perspectives potentielles sur la façon dont les systèmes de réseaux neuronaux artificiels peuvent être améliorés, mais aurait également un impact majeur sur les applications cliniques telles que les interfaces cerveau-machine, les organes sensoriels artificiels (par exemple, les implants de cochlée) ou les prothèses robotiques qui communiquent directement avec le système nerveux. Il a été démontré à maintes reprises que les oscillations neuronales - activité neuronale synchronisée de manière rythmique - jouent un rôle important dans le traitement des signaux corticaux. Cependant, leur rôle exact par rapport au traitement de l'information en cours reste encore très flou. Un problème majeur est que les oscillations neuronales et les modèles de calcul cortical respectifs font des prédictions au niveau des couches corticales. Bien que relativement facile d'accès dans les modèles animaux, l'activité corticale avec une résolution au niveau laminaire ($< 1 \text{ mm}$) est difficile à obtenir chez les sujets sains. La présente thèse vise non seulement à démontrer comment combler le fossé entre les modèles humains et animaux d'un point de vue méthodologique en mettant en œuvre des pipelines de neuro-imagerie de pointe, mais aussi à cibler les prédictions fondamentales dérivées des modèles animaux et des études sur les patients en ce qui concerne le rôle des oscillations neuronales dans les calculs locaux et la dynamique des réseaux cérébraux. Une expérience EEG-IRMf simultanée a été menée avec une taille de voxel de $0,8 \text{ mm}$. Il a ainsi été démontré que l'activité α spécifique aux caractéristiques du stimulus répondait de manière différentielle aux stimuli préférés des neurones activés par rapport aux stimuli non préférés par ces neurones, principalement dans les couches corticales profondes et moyennes de V1, séparément pour les oscillations α de fréquence élevée et basse. On a constaté que les oscillations de la bande γ étaient corrélées au traitement des caractéristiques du stimulus, principalement dans les couches profondes et superficielles. Ces deux résultats sont en accord avec la littérature antérieure. En outre, deux expériences MEG au niveau laminaire ont été menées afin d'affiner les résultats respectifs de l'expérience précédente. Ainsi, une expérience MEG a été spécifiquement conçue pour cibler les interactions corticales de bas niveau, tandis que la seconde a été conçue pour cibler la communication de bas à haut niveau pour des bandes de fréquences multiples. En raison de la pandémie, les enregistrements des données MEG ont été retardés de plus d'un an. Pour cette raison, seuls des résultats préliminaires peuvent être présentés. En outre, il a été prévu d'étudier différentes possibilités de tester les hypothèses respectives à l'aide de simulations de réseaux neuronaux. Cependant, comme prévu, seule la phase initiale de planification et de visualisation a été atteinte. En

résumé, cette thèse fournit de nouvelles informations sur l'activité différentielle des bandes α et γ dans différentes couches corticales, confirmant les prédictions des modèles animaux et théoriques.

Deutsche Zusammenfassung

Das menschliche Gehirn ist eine sehr kostspielige informationsverarbeitende Maschine, die im Vergleich zu anderen Organen relativ viel Energie verbraucht. Diese Kosten müssen durch den potenziellen Nutzen aufgewogen werden, um eine praktikable Lösung darzustellen. Daher lässt sich die Hypothese aufstellen, dass das menschliche Gehirn ein äußerst effizienter Informationsverarbeitungsapparat ist, der jedes künstliche System in Bezug auf die Energieeffizienz bei weitem übertrifft. Das Verständnis der allgemeinen zugrundeliegenden Rechenprinzipien ermöglicht nicht nur neue Erkenntnisse darüber, wie künstliche neuronale Netzsysteme verbessert werden können, sondern hätte auch große Auswirkungen auf klinische Anwendungen wie Gehirn-Computer-Schnittstellen (BCI), künstliche Organe (z. B. Cochlea-Implantate) oder Prothesen, die direkt mit dem Nervensystem kommunizieren. Dass neuronale Oszillationen - rhythmisch synchronisierte neuronale Aktivität - eine wichtige Rolle bei der kortikalen Signalverarbeitung spielen, wurde bereits unzählige Male nachgewiesen. Ihre genaue Rolle im Hinblick auf die tatsächlich ablaufende Informationsverarbeitung ist jedoch noch weitgehend unklar. Ein Hauptproblem dabei ist, dass neuronale Oszillationen und entsprechende kortikale Berechnungsmodelle Vorhersagen auf der Ebene der kortikalen Schichten machen. Während die kortikale Aktivität in Tiermodellen relativ leicht zugänglich ist, ist sie bei gesunden menschlichen Probanden nur schwer mit einer Auflösung auf Schichtebene ($< 1 \text{ mm}$) zu erfassen. Die vorliegende Arbeit zielt nicht nur darauf ab zu zeigen, wie die Lücke zwischen Mensch und Tiermodell methodisch überbrückt werden kann, indem modernste Neuro-Imaging-Verfahren angewendet werden, sondern zielt darüber hinaus auf zentrale Vorhersagen ab, die aus Tiermodellen und Theorien zur Rolle neuronalen Oszillationen abgeleitet wurden. Es wurde ein simultanes EEG-fMRI-Experiment mit einer Voxelgröße von $0,8 \text{ mm}$ durchgeführt. Dabei konnte gezeigt werden, dass stimulusmerkmalspezifische α -Aktivität auf bevorzugte im Vergleich zu nicht bevorzugten Stimuli vor allem in tiefen und mittleren kortikalen Schichten von V1 differenziert reagiert, und zwar getrennt für niedrige und hohe α Frequenzen. Beide Ergebnisse stehen im Einklang mit der bisherigen Literatur. Zusätzlich wurden zwei MEG-Experimente auf Schichtebene durchgeführt, um die entsprechenden Ergebnisse aus dem vorherigen Experiment zu verfeinern. Dabei wurde ein MEG-Experiment speziell auf kortikale Interaktionen auf niedriger Ebene, und das zweite Experiment auf die Kommunikation niedriger bis höherer Ordnung in mehreren Frequenzbändern, ausgerichtet. Aufgrund der Pandemie und anderer Umstände haben sich die Erhebungen der MEG-Daten um mehr als ein Jahr verzögert. Aus diesem Grund können nur vorläufige Ergebnisse vorgelegt werden. Zusätzlich ist geplant, verschiedene Möglichkeiten zu untersuchen, um entsprechende Hypothesen mit Hilfe von Simulationen mit neuronalen Netzen zu testen. Wie zu erwarten war, wurde jedoch nur die erste Planungs- und Pilotphase erreicht. Zusammenfassend lässt sich sagen, dass diese Arbeit neue Erkenntnisse über die differentielle α - und γ -Bandaktivität in verschiedenen kortikalen Schichten liefert und die Vorhersagen aus Tier- und Theoriemodellen bestätigt.

Contents

1	Introduction	1
1.1	The human brain: highly efficient, challenging to study	3
1.2	Structural aspects of the brain	6
1.3	Temporal aspects of the brain	21
1.4	Research question	45
2	Experimental work	57
2.1	COVID 19 remark	57
2.2	EEG-fMRI experiment	58
	Feature specific neuronal oscillations in cortical layers (in prep)	58
	Auto-regressive rank order similarity (aros) test (pre-print)	96
	Software developments for the EEG-fMRI experiment	111
2.3	MEG experiments	117
2.4	Artificial neural networks	151
3	General Discussion	161
3.1	Summary of results	161
3.2	Interpretation in the light of current theoretical frameworks	163
3.3	Overall progress of the project	168
3.4	Closing remark	170
3.5	Evaluation of Hypotheses	171
4	My personal framework: The convolutional brain	173

Acronyms

ACC anterior cingulate cortex	150
AI artificial intelligence	3
ANN artificial neural network	3
ANOVA analysis of variance	143
BCI brain computer interface	19
BIDS Brain Imaging Data Structure	114
BOLD blood oxygenation level dependent	6
CNN convolutional neural network	17
CNO communication via nested oscillations	45
CNR contrast-to-noise ratio	72
CSD current source density	13
CSF cerebrospinal fluid	71
CTC communication through coherence	37
DCM dynamic causal model	24
DICS dynamic imaging of coherent sources	143
DNN deep neural network	10
DOI digital object identifier	114
dVis dorsal visual cortex	52

ECoG electrocorticography	25
EEG electroencephalography	5
EPI echo planar imaging	61
EPSP excitatory post-synaptic potential	20
ERD event-related de-synchronization	27
ERF event-related field	24
ERP event-related potentials	24
ERS event-related synchronization	28
FA false alarm	96
FEF frontal eye field	10
FFA fusiform face area	17
FFT fast Fourier transform	26
fMRI functional Magnetic Resonance Imaging	6
GAN generative adversarial network	7
GBI gating by inhibition	35
GLM general linear model	82
IAT implicit association task	140
IFJ inferior frontal junction	38
IPSP inhibitory post-synaptic potential	20
ISI inter stimulus interval	64
L1 layer 1	11
L2/3 layer 2/3	11
L4 layer 4	11

L5/6 layer 5/6	11
LCMV linearly constrained minimum variance	69
LFP local field potentials	13
LGN lateral geniculate nucleus	10
LIF leaky integrate and fire	16
LMF left medio-frontal	144
LOT left occipito-temporal	144
LSTM long short-term memory	17
MCP multiple comparison problem	96
MEG magnetoencephalography	5
MRI magnetic resonance imaging	45
MT middle temporal	16
MTG medial temporal gyrus	52
MUA multi unit activity	30
PET positron emission tomography	23
PFC pre-frontal cortex	11
PPA parahippocampal place area	38
pRF population receptive field	6
ROI region of interest	54
RT reaction time	135
rTMS repetitive transcranial magnetic stimulation	28
SD standard deviation	72
SNN spiking neural network	16

SNO spiking neural oscillator 155

SNR signal-to-noise ratio 30

SOA stimulus onset asynchrony 32

SSVEP steady state visual evoked potential 129

TE echo time 61

TEO temporo-occipital 32

TR repetition time 61

VAN ventral attention network 149

vVis ventral visual cortex 52

VWFA visual word form area 51

1 Introduction

Plants do neither have a brain nor a nervous system [2], which is surprising, given that they do communicate [3], adapt to environmental changes [4] and engage in quasi-social interactions [5]. These seemingly complex patterns of behavior are classically reserved for or attributed to organisms possessing nervous systems. According to Watanabe et al. (2009) [6], "the function of the nervous system is to sense and relay fast information about surroundings" and further, "this rapid and restricted mode of signal transmission allows an animal to process multiple messages and respond appropriately". According to this definition, the purpose of a nervous system within an organism is limited to functions that even organisms without such a complex system can accomplish, except for the two main keywords: *process* and *fast*. Those two keywords however are of major importance for the understanding of complex systems, such as the human brain. *Processing* information can be interpreted synonymous to *changing* information. In biological systems the change of information might often be an actual *reduction* of information, such that inputs from the environment can be meaningfully interpreted and reacted to. An extreme example would be the occurrence of a predator in the field of view of some animal. The electromagnetic frequency spectrum perceived as light of different color by the retina or complex short lived air pressure changes in form of auditory wave patterns, are reduced to the information of the presence of a predator in order to respond appropriately (e.g. fight or flight). Hence, (complex) information processing is of increased relevance if a specific input cannot be relayed one-to-one onto a specific output (like e.g. a simple chemical reaction). Acacia trees for instance warn each other by releasing molecules into the air if eaten by a herbivore. Neighboring trees receive odor-like molecules at specific receptors and a short cascade of chemical reactions initiates a chemical defense program [7]. Importantly, there is no ambiguity within the signal cascade. The signal molecule directly causes the appropriate response without changing (processing) the information as such. Receiving the signalling molecule and initiating production and release of defense molecules does not require the *change* of information and rather acts similar to a light switch. Hence, an organism's behavior relying on simple signal transmission cascades is limited to a predetermined number of input-output patterns. Adaptation to environmental changes are hence slow or even only possible on an evolutionary time scale. Neuronal information processing enables emergent¹ behavior that is less determined by the physical substrate of the organism. *Fast* responses enable an organism not only to react

¹Emergence results from the interaction of different parts of a particular system that cannot be explained by adding up the features of the respective parts [8]. Similarly, the interactions of atoms within a tree or a human can have dramatic consequences in what emergent behavior they allow for. Life itself is based on "dead" matter, which by interaction produces emergent features. Emergent properties have been observed in biological signaling pathways as well [9]. Bi-stable network state patterns (e.g. steady-state activities) or timed signal integration has been observed (among other patterns) to "emerge" depending on input properties.

quickly to changes of the environment, but furthermore to manipulate the environment and receive (almost) instant feedback. Especially the latter aspect seems of major importance, since nervous systems can be almost exclusively observed in moving animals. Famously, *juvenile sea squirts* travel the ocean in search for a place to cling to for the rest of their lives and once they found it, start consuming their own brain, because its no longer needed. Without the need for rapid information processing and adaptation, those animals decide to get rid of such a costly and otherwise not vital system. Any understanding of functional principles for any kind of biological nervous system, can hence partly be derived from the functional purpose it is serving. In other words: Computational properties in the brain are tightly linked to the respective purpose they serve. A *mantis shrimp's* compound eye for instance was found to contain an array of 16 different photo-receptors for sensitivity to linear and circular light polarization as well as a multitude of different wavelengths [10]. Exaggeratedly spoken, those animals have 16 base colors (as opposed to three in humans). Those findings have been explained, by the predatory nature of those animals and their capability to communicate in the short wave ultra violet light spectrum with others, unseen from most - if not all - other marine animals. The behavior of the *mantis shrimp* is heavily driven by visual input - that is the sensing and interpretation of a specific spectrum of electromagnetic radiation - and outperforming other marine animals at the level of electromagnetic spectral bandwidth provides a potential evolutionary leap. Evolutionary advantages however, do not necessarily mean that "the best solution" for any given problem was found, but rather that the solution that was found was sufficient enough to minimize *that* specific disadvantage that otherwise would have led to extinction. If plants do not possess a nervous system, this means that this system is not crucial to their evolutionary niche and respective "behavioral" patterns. Similar to the concept of "embodied cognition" ² [11], environmental conditions and evolutionary niche seem tightly linked to the type and capacity of biological nervous systems. The human pre-frontal cortex has been found to be particularly large, compared to chimpanzees and macaques [12] and has been linked to cognitive control [13], a psychological term describing the underlying effort tied to problem solving, working memory, etc., which could be considered core evolutionary features in humans.

²Embodied cognition [11] reflects the idea that an individuals perception of the world is heavily determined by the shape and physical properties of its body. A very simple example would be the taste of sugar. Because glucose is one of the most fundamental energy sources for life in general its importance is cognitively reflected by a pleasant taste (sweet). Another examples would be sheer body size. Depending on the size of an organisms body, aspects of the environment that are cognitively classified as "small" or "large" might heavily vary.

1.1 The human brain: highly efficient, challenging to study

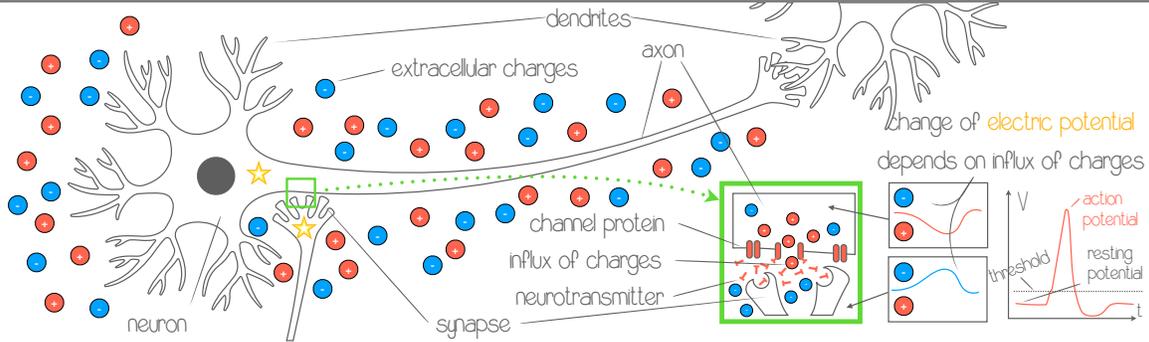
The human brain is a remarkably complex and efficient information processing machinery. Complex, because even letting temporal dynamics aside, it contains close to 100 billion neurons [14] and an estimated number of about 100 trillion connections [15]. Efficient, because even though it consumes almost 20% of the energy provided to the body while only weighting in at 2% of its mass [16], its total power consumption is estimated to be around 20 *Watt/h* [17], which is roughly the same consumption as a modern LED-light-bulb. It has been estimated that a real time simulation of the human brain on standard compute hardware, would result in a power uptake in the *Gigawatt/h* range [18]. Even though those estimates must be seen as highly speculative, a difference in power consumption between regular transistor based computers and the human brain very likely spans multiple orders of magnitude. Understanding how computations are carried out by the human brain therefore might provide important insight into efficient and adaptive computing in general. While the human brain can be considered rather weak in the computation of exact solutions (e.g. mental arithmetic), approximate solutions in turn are computed very rapid and efficiently (e.g. detecting and responding to an open door of a car parking next to the cycling lane). Modern applications in artificial intelligence (AI), employing artificial neural network (ANN) models, could greatly benefit from insights into the efficient compute machinery, its core functions and architecture. Beside the rather distant goal of developing new methods for information processing derived from biological systems, insights into computational patterns play an important role in the understanding of humankind in itself, as well as for the development of treatments for brain related disease. While the first can be seen as a more or less philosophical issue, the latter directly impacts the lives of people suffering from brain related disabilities. Understanding the anatomical and functional properties of the human brain has thereby led to the development of deep brain stimulation to ease symptoms related to Parkinson's disease [19], made retina [20] and cochlea implants [21] feasible, as well as brain controlled prosthetic limbs [22]. Beside the major challenge of precisely mapping out anatomical connections in the order of trillions, a full understanding of human brain functions must entail a fundamental understanding of the temporal dynamics of neuronal communication and respective computational properties. Understanding the temporal dynamics requires an understanding of the underlying "hardware", which can be achieved on multiple scales (for a short summary about single cell neuronal activity see WHAT IS ... on page 5). However, this condition is not sufficient, as the example of the *Hydra* demonstrates. The *Hydra* is a small freshwater organism with a remarkably simple neuronal network. Its nervous system is composed of only a few hundreds to few thousands of neurons [23], which allowed researchers to create the first fully complete neuronal activation map of a living organism, by recording the activity of all neurons simultaneously [24]. Its nervous system - due to its simplicity called nerve net - contains only two sheets or layers of neurons [25] that are functionally largely separated [24]. One of those sheets mostly comprises sensory neurons (receptors), while the second

sheet coordinates behavior, such as feeding, elongation, radial and longitudinal contractions. Notably, the *interaction* between anatomical structure and temporal dynamics led to rather complex patterns of emergent behavior in the *Hydra* model system. Interestingly, even for such a simple model organism as the *Hydra*, it has not yet been possible to decipher the full neuronal code, since the combinatorial explosion for different states of activation for each neuron, together with the complexity added by temporal dynamics across the network(s), creates a huge state space at each time point. For much more complex neuronal systems as the human brain, even obtaining a full activation map might be an intractable task for the foreseeable future. Focusing macroscopically on functionally distinct compartments (such as the visual system in mammals) instead of the entire brain including each single connection and activation state however, still provides meaningful insights into some computational core principles of the mammal cortex (similar to unraveling the two functionally distinct networks in the *Hydra*).

Studying brain network dynamics on a macroscopic scale (not looking at individual neurons) can provide highly accurate insights about the underlying functionality, as demonstrated in simulation studies on mean field or neural mass models [26–28]. The estimated temporal dynamics produced by a simulated spiking neural network (simulating the dynamics and interaction of each single neuron) thereby can often be described very precisely from global dynamics obtained from the entire network and in turn allows for precise models using mean field or neural mass dynamics without the need for single neuron simulations. In the context of human brain research, those results suggest that meaningful information about the fundamental principles of brain network dynamics might be at least partly extractable on a macroscopic scale. Nonetheless, synaptic density e.g. in the adult human prefrontal cortex is estimated to lie in the range of 10^9 synapses per mm^3 [29]. Recording the average activity of this many interactions per unit of time limits the accuracy of macroscopic information that can be extracted nonetheless, almost irrespective of model accuracy. Therefore, even on the macroscopic scale, a precise description of anatomical conditions *and* temporal dynamics are indispensable to formulate accurate hypothesis to explain (human) brain functions.

Due to the tremendous amount of functions that are carried out by the human brain, a full functional overview as well as an in depth comparison of anatomical differences between multiple cortical and sub-cortical areas and functions cannot be provided within the realm of this work. Instead this work focuses on the human neo-cortex as a structure of interest and will specifically investigate the visual system as a model system for perception and sensory information processing in general. Thereby, anatomical and functional properties will be related to current frameworks on local cortical computations as well as widespread network dynamics. Structural aspects of the brain will be discussed in the light of Temporal aspects of the brain approaching the Research question, that is the role of neuronal oscillations in local computations and network brain dynamics.

What is ... neuronal activity (electrically speaking)?



Neurons are single cells with long extensions and branched out structures (axons and dendrites) which form the fundamental units of information processing in the brain. Inside the neuron - and outside in the extracellular space - a variety of ions, such as positively charged (\oplus) sodium (Na^+) or potassium (K^+) and negatively charged (\ominus) chloride (Cl^-) and other organic anions can be found in different concentrations. This causes an electro-chemical gradient and an electric potential across the neuron's membrane of $\approx -70\text{ mV}$. Neuronal information transfer is initiated by the pre-synaptic cell releasing neurotransmitters into the synaptic cleft, which causes a conformational change of channel proteins of the post-synaptic cell. Ions from the extracellular space can now pass the cell membrane of the post-synaptic neuron. This influx is driven by diffusion ("chemical gradient") and the difference in electric potential ("electric gradient"). Voltage gated channel proteins within the post-synaptic neuron's membrane perform a conformational change once a specific voltage threshold is hit. Now even more ions can flow in, which in turn causes a sharp peak in the membrane potential (action potential). Once an action potential is released, the electro-chemical gradient "travels" along the axon, where eventually neurotransmitters are released into the cleft between the current post-synaptic and the next post-synaptic neuron. In case of an inhibitory incoming signal, negative charges flow inside the neuron, lowering the resting potential and decreasing the probability to release an action potential even further. The combined potential changes caused by thousands of synapses that arrive at varying parts of the neuron are integrated at the membrane and only if the combined signal raises the membrane potential sufficiently, an action potential is released. The exchange of ions with the extracellular space, as well as the resulting current flow inside and outside the neuron causes electromagnetic field changes, which - summed up over a large number of neurons - can be measured even from outside the head (see WHAT IS ... electroencephalography (EEG) or magnetoencephalography (MEG) on page 20 or page 22). This info-box was first referenced on page 3.

1.2 Structural aspects of the brain

The human cortex: a modular, multipurpose neural network

Neurons form the fundamental building blocks of any neural network (hence the name). Those cells are highly interconnected with each other via axons and dendrites, receive a weighted input and produce an output of a certain strength that serves as an input for other cells. An immense microscopic machinery thereby ensures proper functionality of the cells and modulates cellular activity. This includes neurotransmitter production and transport to synapses, the regulation of synaptic growth or decay as well as the generation of a resting state electro-chemical gradient across the cell membrane with a charge of around $-70mV$. Irrespective of the immense amount of intra-cellular regulatory processes, inter-cellular connections determine to a large degree the respective functionality of a neuronal assembly (i.e. local compute module). Famously, Hubel and Wiesel (1962) [30] identified cell populations in cat's primary visual cortices, responding to a specific orientation and / or movement direction of high contrast bars projected onto the retina. While bars of specific orientation have been presented to the sedated cats, neuronal response patterns were recorded using micro-electrodes from multiple single cells of the cat cortices. Depending on the orientation of the bar, or its respective movement direction, different cells increased or decreased their firing rate differentially. It has been hypothesized that cells specifically responding to bright and cells responding specifically to dark areas, are interconnected, such that the combined neuronal population responds to a specific orientation of a bar. In this model, the edge of a bar would be detected, if a cell responding to the dark area of a bar and a neighboring cell responding to bright areas (no bar), fire simultaneously [31]. If specific bar orientations are coded within the cortical network structure itself, then the hypothesis of a correspondence between retinal and cortical activation - as some form of direct mapping - follows consequently. Indeed, studies employing functional Magnetic Resonance Imaging (fMRI) (see WHAT IS ... on page 8 for general information about fMRI) revealed a direct link between the retinal input image and the blood oxygenation level dependent (BOLD) signal (which serves as a proxy for neuronal activation [32]). Specific locations of the retinal input image were found to map to specific locations of the visual cortex, preserving spatial relationships to a large degree. The method of population receptive field (pRF) mapping can be used to an extent where it is possible to reconstruct visual input stimuli, based on the spatially distributed BOLD signal in primary visual regions [33]. Interestingly, similar activation patterns (i.e. stimulus reconstructions from BOLD) could be observed in participants that were only imagining the respective shapes. Furthermore, it could be demonstrated that even illusory shapes (e.g. the triangular shape produced by the Kanizsa illusion ) , yield spatially correlated brain activation patterns corresponding to the presumed physical (but illusory) shape [34]. Nonetheless, those retinal-cortical mappings cannot be described as 1 : 1 mappings (e.g. transferring "pixel" locations correspondingly). Boskin et al. (1997)

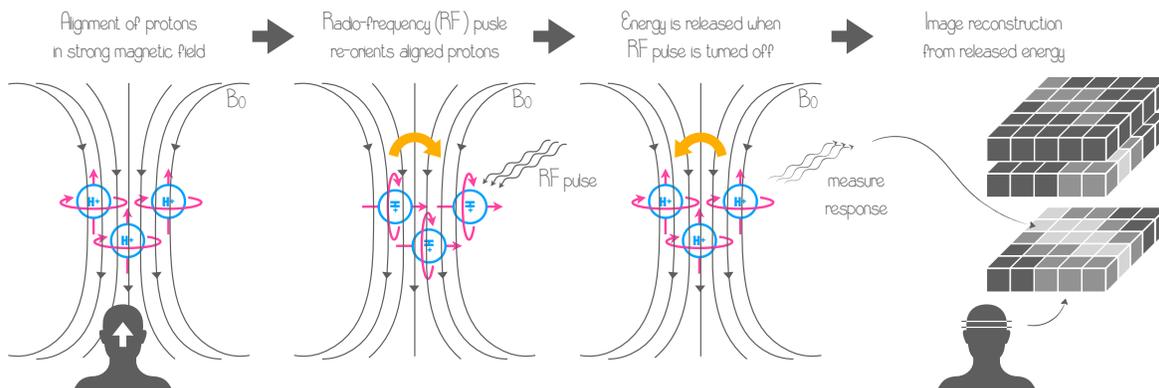
[35] investigated orientation specificity in V1 of tree shrew and systematically mapped the orientation of stimuli to locations on the cortical surface sheet. While not random and locally highly correlated, the resulting map does not follow any linear pattern. It has been hypothesized that orientation maps of the primary visual cortex are the result of self organization during development [36], based on the organization principle of Kohonen maps³ [37]. The fact that there is no 1 : 1 mapping between retinal stimulation and cortical representation, but nevertheless a non-linear (traceable) mapping means that it could be described by an unknown function, which potentially could be approximated by some sort of machine learning algorithm. Previous research demonstrated that using a variety of statistical and machine learning techniques, such as gaussian mixture models or by employing a generative adversarial network (GAN), stimuli can be reconstructed with a high degree of accuracy from cortical activity, represented in the BOLD signal [38–40].

The human sensory cortex: mapping ND to 2D

A similar (locally correlated but non-linear) mapping between real world and cortical representations can be observed for the cortical homunculus, where adjacent body parts are represented adjacently on the surface of sensory and motor cortices [41]. Furthermore, a relationship between physically correlated input signals and spatially correlated cortical representations can be observed in the auditory domain. It has been found that similar frequencies of an auditory signal can be tonotopically mapped to neighboring cortical locations within the primary auditory cortex [42, 43]. Not surprisingly - though harder to map - specific spatial organizations have even been observed in the olfactory domain. In mice, the similarity of odors of specific molecular structures, could be successfully mapped to spatially correlated locations within the olfactory bulb of mice [44]. Whether this observation can be extrapolated to human brains remains to be investigated but, given the large similarity between sensory mappings across mammal cortices, it can be considered likely. A major challenge for odor related research in general, is to classify "adjacency" in the physical world. Nevertheless, qualitative proximity of a stimulus within the perceptual space, appears to be tightly linked to structural proximity within the mammal cortex in any (or at least many) domain(s). A large part of sensory processing within the human brain can hence be explained by its three dimensional structure (in the examples the two dimensional projection to the cortical surface which is of course a three dimensional physical structure).

³Kohonen maps [37] are produced by artificial neuronal networks. High dimensional patterns are mapped to a discrete bi-dimensional feature space via unsupervised self organization. This type of dimensionality reduction happens e.g. when an object drops a shadow by obscuring a light source. Thereby, the shadow is an accurate mapping of a three dimensional object to a two dimensional plane. The fundamental idea has been derived from observations about the cortex. It has e.g. been found that specific bar locations and orientations are mapped spatially correlated to the approximately two dimensional surface of the cortex, where spatial and orientation specific proximity determines the respective location in the visual cortex with some sort of non-linear - but locally correlated - mapping [35].

What is ... functional Magnetic Resonance Imaging (fMRI)?



Far more than half of all atoms in the human body are hydrogen. They consist of only a single proton that can easily be influenced by an external magnetic field, because due to its spin the proton has a magnetic moment. A strong magnetic field (B_0), causes all hydrogen cores in a region to align along the magnetic field lines. Using a radio-frequency (RF) pulse, the aligned protons are collectively realigned relative to B_0 (e.g. 90° or 180°). After switching of the RF coils, hydrogen atoms realign again with B_0 and thereby release the energy that they previously absorbed from the RF pulses. Depending on the tissue composition (e.g. how much hydrogen and in which form it is bound), absorption and release energies differ, which can be mapped to different brightness values of the resulting image. By probing multiple spatial locations (recording the signal and reconstructing the respective images) respective MR images are created slice wise. For this to work, the RF pulse needs to precisely hit the resonance frequency of the target (e.g. proton). However, this also enables probing of different atoms or molecules. It has been discovered that oxygenated and de-oxygenated hemoglobin expose a different magnetic behavior [45]. Activated neurons consume glucose and oxygen at a higher rate (to synthesize ATP) causing a hemodynamic response (increased blood flow to active regions) to compensate for the increased uptake. Thereby, the ratio between oxygenated and de-oxygenated hemoglobin changes (and hence the magnetic response), which can be detected using fMRI. This is why the resulting signal is called BOLD signal. Since the hemodynamic response is slow (peak after ≈ 6 s) fMRI has a poor time resolution. In turn the spatial resolution can be even lower than 1 mm. This info-box was first referenced on page 6.

However, this direct mapping can only be observed for primary sensory regions and does not extrapolate well to higher order association areas of the brain. Experiments investigating e.g. inattentive blindness demonstrate however that even fully visible objects (i.e.

objects that are clearly represented on the level of the retina), are not consciously perceived, if specific conditions are met (or not met) [46]. A retinal representation, followed by primary cortical processing, is hence not a sufficient condition for incoming visual stimuli to be actively perceived. Higher order brain regions mediate or even suppress early sensory processing across all aforementioned domains [47–49]. These observations led to the insight that brain regions cannot be described as functionally distinct units that compute an output that is sent to higher order regions where it is processed further, but rather that the brain acts as a highly interconnected system, where even primary sensory regions are embedded in a large modular network structure [50]. Thereby feed-forward and feedback connections can be found locally within sensory regions [51–55] as well as long-ranging connections between vastly distinct cortical areas [56–58]. Notable, even though primary sensory and somatosensory regions follow the principle of similarity in the physical space that is represented in spatial similarity in the cortical space, the influence of higher order regions on primary regions, as well as how primary sensory information is integrated with higher order brain functions, remains to be understood. Even though the human cortex is a highly complex structure, its repetitive, modular organization [50] allows potentially to transfer specific core computational principles between sensory modalities.

A second core principle that exposes great extrapolation value is *convergence* [59] (see Figure 1 lower left). Convergence in the given context means that during each hierarchical processing step, the dimensionality of the input data gets reduced. This is achieved by grouping the signal of many neurons. Anatomically this means that many neurons are connected to a few from one hierarchical level to the next. An early example of cortical convergence are the findings by Hubel and Wiesel (1962) [30], where simple cells (responding to light ON or OFF at specific spatial locations) converge to complex cells (that receive input from many simple cells) that in turn abstract simple cell respond to e.g. enable spatial in-variance of input features (such that the actual location of a specifically oriented bar does not matter anymore, but only its orientation). Low level features are combined into (more abstract) high level features progressively along the hierarchy exploiting the principle of convergence [59]. In order to integrate new information in its entirety (e.g. multiple sensory information), modality specific analyses in respective sensory cortices extract task relevant features, converging more and more towards higher hierarchical levels until processed signals from multiple modalities again converge into a combined representation [60]. Top-down feedback signals however *diverge* towards low hierarchical levels. Converging feed-forward and diverging feedback connectivity combined with the principle of modularity potentially reflect a core principle that is used throuout the entire cortex.

The human visual cortex: a model system

The visual system has traditionally been *the* model system for studying perception related phenomena in higher order mammals, especially in humans, due to the relative ease at which stimulation can be controlled and performed. Hence, the visual system is probably the most studied perceptual system in humans, but - as described above - could potentially serve as a model system for investigating cortical computational properties in general. Hierarchically organized, visual information is routed from retinal cells via the thalamus to primary visual areas V1-4 [61] (see Figure 1). Retinal ganglion cells, respond specifically to certain wavelength of the electromagnetic wave spectrum [62]. Furthermore, the presence of ON and OFF ganglion cells could be demonstrated in many mammals [63, 64], including primates [65, 66], including humans [67]. The combination of cells responding to the presence of light (ON cells) or its absence (OFF cells), form the basic neuronal coded visual input to the brain. In a first step, signals from retinal ganglion cells are routed through the lateral geniculate nucleus (LGN) of the thalamus, with feed-forward connections to primary visual area V1 that resemble similar receptive field shapes and sizes as found on the retinal level [68]. Interestingly, those afferent cells from the retina to the LGN, comprise only $\approx 10\%$ of input to LGN relay cells, whereas $\approx 30\%$ stem from feedback connections from V1 [69]. Similar to thalamo-cortical feed-forward connections, feedback connections from V1 to the LGN follow the principle of retinotopic organization. It has been hypothesized that feedback connectivity to the thalamus enables sharpening of the receptive field and enhancing signal transmission of activated receptive field neurons [70]. Even though non-linear mappings of receptive field specificity could even be found in frontal areas of the human cortex [71], more direct mappings between the visual input signal and spatially distributed cortical responses are mainly found for areas V1-4 [72]. Thereby, similarities between the hierarchical organization of the brain and the structure of a deep neural network (DNN) ⁴ can be observed [74] (see Artificial neural networks: Not so distant relatives). On a fundamental level, elementary image features, such as lines are processed early in the visual stream (V1) [30] and are successively combined to form higher order features - such as edges in V2, higher up in the hierarchy [75]. Additionally, V4 plays a critical role in gating information by attentive processes. In macaque monkeys, electric stimulation of specific neurons in the frontal eye field (FEF) led to a modulation of the signal recorded in V4 [76]. Stimulation of retinotopically corresponding areas in the FEF led to an increased response, whereas stimulation of not cor-

⁴Deep neural networks (DNNs) are a class of machine learning algorithms that have been inspired by the cortical architecture of mammals [73]. For a detailed discussion see "Artificial neural networks: Not so distant relatives". In a nutshell, a simple DNN consists of an input, an output and N hidden layers. In each layer the output is computed from the weighted input. The term deep can be derived from the amount of hidden layers. During training the weight matrix of each layer (*input* \times *output*) is modified, such that a specific input leads to a particular output. Since only little insight into the inherent "logic" behind a trained network can be gained, layers between input and output are called hidden and the network in general is considered a "black box" with respect to its trained "decision tree".

responding neurons led to a suppression of respective V4 neurons. It has been hypothesized that these results mimic the process of the presence or absence of covert or overt attention to stimuli [77]. However, long-range connections from other sensory areas as well as higher order cortical regions, such as pre-frontal cortex (PFC) to V1 have been reported as well [78]. Furthermore, it could be demonstrated that feedback connections from V4 to V1 [52] modulate incoming visual signals at the earliest stage of the cortical visual processing hierarchy [79]. The fundamental architecture of the early visual processing pipeline (retina \rightarrow thalamus \rightleftharpoons primary visual cortex) can hence be described as a hierarchical neural network with feed-forward and feedback connections that utilizes hierarchical feature processing to efficiently extract visual information in a bottom-up / top-down feedback-loop.

On the local level of microcircuit connections within primary visual regions, feed-forward and feedback connections are routed through different cortical layers depending on the processing hierarchy [80] (see "Canonical microcircuits: cortical compute modules"). The percentage of neurons in each area that fall into supra-granular layers thereby reliably predicts the rank in the processing hierarchy in the visual system [81]. Incoming signals arrive at layer 4 (L4) - either from thalamic connections or lower level cortical regions - and are internally routed to layer 2/3 (L2/3), where feed-forward connections to higher order regions and to layer 5/6 (L5/6) within the same region are found [82]. Intra-areal feedback from L2/3 to L5/6 is complemented by feedback connections from higher order regions, such that higher order layer 1 (L1) and L2/3 and L5/6 are connected in feedback direction to their lower order counterparts. Additional intrinsic feedback connections between L5/6 and granular layers (L4) have been identified as well [83] (see Figure 1). In turn, the feed-forward flow of information originates mainly in thalamic nuclei (LGN), superficial cortical layers (L2/3) or cortical layer 5 [80] and is routed to L4 (from L2/3) or layer 5 of the target region. It has been hypothesized that dendritic cortical microcircuits are related to error back-propagation - an algorithm used to train deep neural networks [84] - by exploiting feedback directed prediction error signals [85]. Even though classic error back-propagation as used in deep learning has been deemed neuro-biologically implausible [86, 87], more recent hypotheses indicate that error feedback might be implemented different than classical back-propagation, but nevertheless exposes functional similarities that are represented by the cortical microcircuitry as well [88]. Whether the proposed architecture of canonical microcircuits extends to cortical areas other than striate cortex is still under debate, since the cortical architecture varies strongly depending on the precise location [89, 90]. Fundamental principles about the cortical architecture however can be identified irrespective of the stimulus modality: a general hierarchical [80, 81, 91], converging and diverging [59, 60] modular network organization [50]; anatomical and functional distinct layers [52, 80, 82]; bi-directional information flow [57, 82, 91] and spatial proximity in the cortex for representational proximity of stimulus features [30, 41, 43]. Computational and theoretical models about the brain's compute architecture must necessarily comprise a model for canonical microcircuits *and* functional (long)

range connectivity (see "Canonical microcircuits: cortical compute modules" and "Functional and effective connectivity"), because local computations *and* interactions between different network parts together compute a coherent percept of the environment.

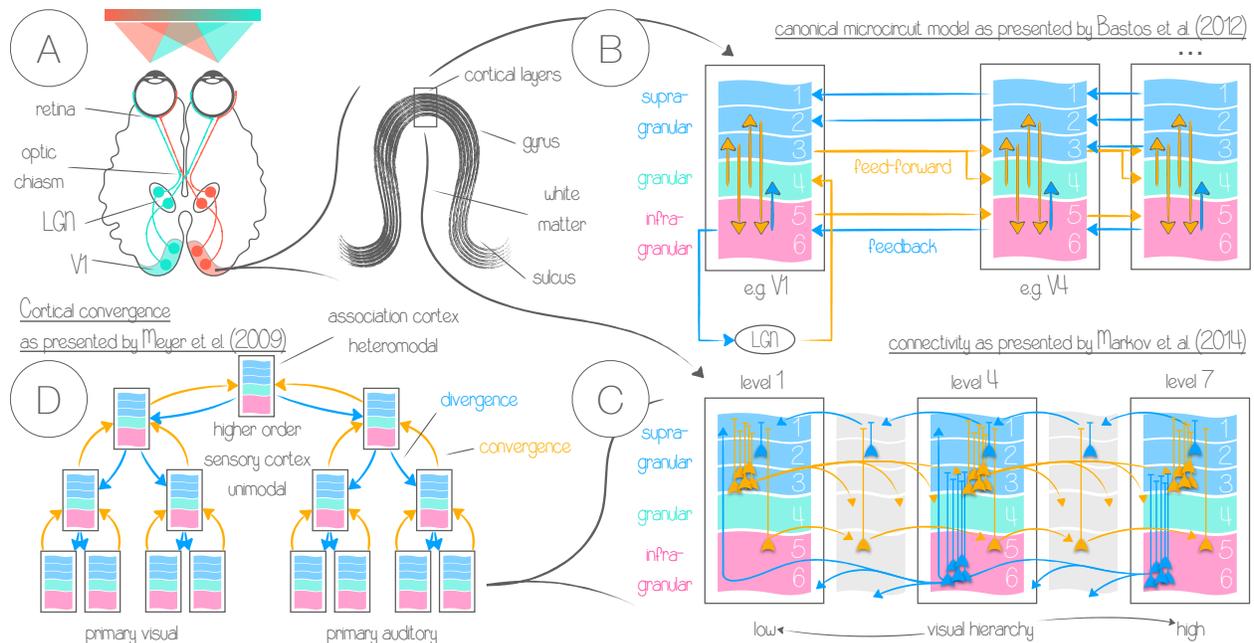


Figure 1: Illustration of fundamental visual pathways, canonical microcircuits (see "Canonical microcircuits: cortical compute modules") and inter-area connectivity. **(A)** Information from the left visual hemifield is routed through respective parts of the retina and the optic nerve, through the optic chiasm to the respective right LGN of the thalamus and from there to primary cortical visual regions (similar for right visual hemifield, but mirrored). Cortical columns consist of a composition of around six cortical layers [89] that expose feed-forward, feedback and intra-area connections that are functionally interlinked to form a canonical microcircuit [82]. **(B)** As presented by Bastos et al. (2012) [82] feed-forward connections terminate at L4 to deliver the input to the circuit. From there, intra-area connections send feed-forward information to L2/3 where it is transferred to L5/6 and to L4 of the higher order region as input. Deep cortical layers in turn send intra-area feedback to L4 and inter-area feedback to L5/6 of lower order regions. Furthermore, superficial layers send feedback information to corresponding layers of the lower order regions (figure adapted). **(C)** Markov et al. (2014) [80] found additional feed-forward connections between regions in layer 5 in macaque monkeys. In addition to that, deep layer feedback are mostly located at layer 6 (figure adapted). **(D)** Both, within and between sensory and higher order association areas a neuronal convergence / divergence pattern can be observed. Meyer and Damasio (2009) [59] argue that bottom up connectivity is highly convergent (more input than output nodes), whereas top down neuronal connectivity is highly divergent (more output than input nodes). While bottom up convergence is thought to reflect the process of feature integration (like multiple edges to form a shape), top down divergence ensures hierarchy level adjusted feedback signals (figure adapted).

Canonical microcircuits: cortical compute modules

In a nutshell, the idea of cortical canonical microcircuits is based on the idea that the (mammalian) cortex can be modelled better as a set of repeating highly connected modules rather than a more or less randomly connected "super-structure" [92]. Those cortical "micro-modules" are thought of as a general principle of how the cortex is set up. The earliest model for a canonical microcircuit was proposed based on findings in the visual cortex of cats by Douglas et al. (1989) [93]. Single cell neuronal activity from cat's striate cortex has been recorded, while simultaneously the optic radiation (fibers from the LGN to the cortex) was stimulated. It was found that thalamic stimulation could not explain the majority of signals that were recorded at cortical electrodes. Instead cortico-cortical connections explained the majority of the data. Additionally, the authors found synaptic delay timings much shorter than excitatory or inhibitory event evolution. It has been hypothesized that temporal dynamics within the cortex cannot be explained by solely relying on synaptic input timings. The authors proposed that cortical dynamics rely on a microcircuit consisting of three different populations. In that model, incoming signals from the thalamus are routed to L2/3 and L5/6 of the cortex and to a GABA-ergic population of cells with inhibitory connections to aforementioned cortical layers. Cortical layers L2/3 and L5/6 are connected to each other and themselves, as well as to GABA-ergic cells via excitatory connections. Testing the model by measuring the pulse response at the recording sites verified those assumptions. Generality of the canonical microcircuit model has been demonstrated by testing the model's predictions on non-visual cortex areas in a different species. Godlove et al. (2014) [94] tested whether non-sensory frontal areas (the supplementary eye field) would expose similar response patterns in macaque monkeys, as would be expected from findings of the cat's striate cortex. Indeed, local field potentials (LFP) and respective current source density (CSD) patterns measured at frontal electrode sites, resembled the predicted pattern when light was flashed onto the monkey's retina. Current sinks were found for middle layers. Furthermore, local recurrent connections across all cortical layers were verified by a sharp increase in activity after stimulation which again was followed by a steep decrease shortly thereafter. However, initial mid-layer activation, was followed by a spread of activity to superficial and deep layers which is not predicted by the original model proposed by Douglas et al. (1989) [93], where deep and superficial layers would receive direct thalamic input. In the updated model, superficial and deep cortical layers only receive feed-forward input from L4 within a region. Irrespective of this minor deviation and given that original results could be verified using a different method (LFP rather than single cell recordings), different cortical areas (supplementary eye field rather than striate cortex) and across different species (macaque rather than cat), the proposed canonical microcircuit model in general provides a sufficient basis for how fundamental cortical response patterns in mammals can be explained. It was found that the neuronal response of macaque FEF - a region related to top down attention control [95] - could be modelled by the canonical microcircuit model proposed by Douglas et al. (1989)

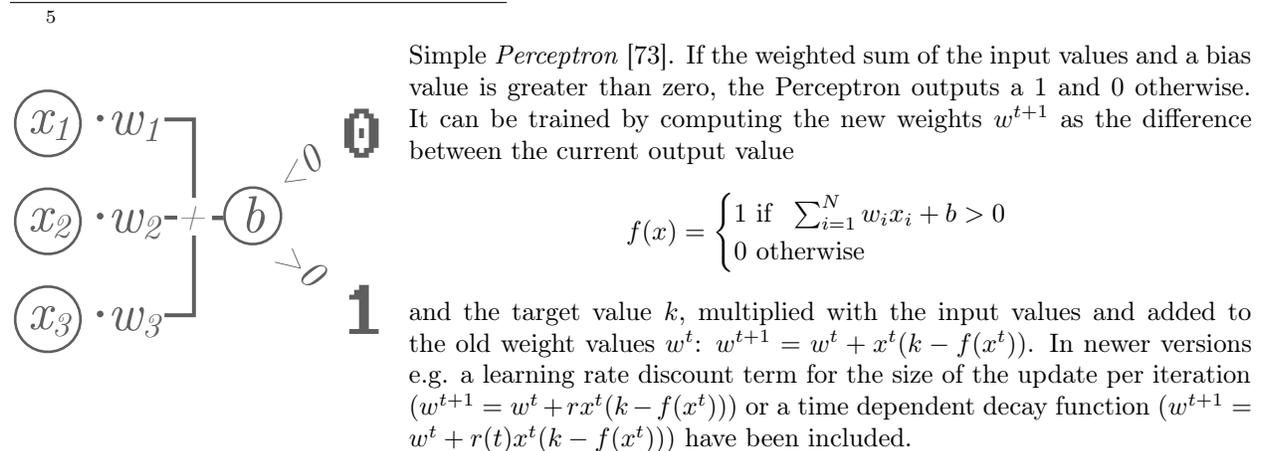
[93] for a variety of tasks performed by the FEF [96]. In an fMRI study by van Dijk and colleagues (2021) [97] numerosity specific neuronal populations in parietal association areas have been identified and investigated in the light of canonical microcircuits. Numerosity specific neurons are tuned to a specific set size (and width) of items. It was found that neuronal responses sharpened (i.e. neurons preferably responding to a given size had lower width of tolerance for other sizes) predominantly in deep and superficial layers. Generalizability of the canonical microcircuit model as a general model for cortical information processing not only follows from numerous findings across multiple cortical areas and species, but is furthermore thought to reflect evolutionary sparsity that enables multiple different functions to be computable on a similar architecture [92]. The canonical microcircuit model thus reflects the idea of *the* fundamental processing unit within the neo-cortex that can be used to compute a multitude of cognitive functions. More complex computations however involve the combination of multiple local networks into spatially spread out dynamic brain networks connected via functional or effective connectivity (see "Functional and effective connectivity"). Since the brain is by definition a neural network and its general architecture includes generality, modularity and modular convergence and divergence, a comparison between the brain and artificial neural networks (especially DNNs) suggests itself. Thereby the interplay between neuroscientific research and developments in the field of artificial intelligence, could mutually benefit. In the following section it will be rolled out, how ANNs can be structurally compared to the (human) brain, where potential comparisons are limited and how insights from e.g. DNN layer activation can help to shed light on the computational principles the brain might rely on.

Artificial neural networks: Not so distant relatives

Artificial neural networks (ANN) as a branch of machine learning have become extremely popular over recent years and made it into technology that is nowadays used by almost every human being. Even though the first implementation of an ANN dates already back to 1957 and was called *Perceptron* [73], only the latest wave of deep learning could make use of new hardware developments and respective efficient implementations, making ANNs tremendously popular during the past decade. The Perceptron was based on theoretical work by McCulloch (1943) [98] and a fully integrated autonomous machine. It comprised a set of input nodes or neurons - arranged in a grid structure - and a single output node. Each input neuron (x_i) was connected via a set of weights (w , with values between -1 and 1) to the output neuron. Before the total weighted sum was computed each input neuron activation value was multiplied with the corresponding weight, before the total (weighted) sum was computed ($\sum_{i=1}^N w_i x_i$). At the output neuron an additional value (bias) was added to modify the neurons responsiveness (similar to the voltage threshold at the membrane of actual neurons). If the computed output value exceeds 0, the Perceptron outputs a 1 and

otherwise a 0⁵. The Perceptron was capable of learning to differentiate between two classes of inputs. After initializing the weights with random values close to zero, for each training iteration (t) weights were updated by simply computing the difference between the target output value (k^t) and the Perceptron's output value ($f(x)$ see Footnote 5) which was then multiplied with the current input activation (x^t) and added to the previous state of the weight vector $w^{t+1} = w^t + x^t(k - f(x^t))$. Both, McCulloch (1943) [98] and Rosenblatt (1957) [73] have explicitly stated that their work resembles biological processes that could be found in the (human) brain. Indeed, at a first glance the resemblance is remarkable: The brain consists of single cells that could either be active (an action potential is released) or not (no action potential is released). Neuronal signals are transferred via synapses, which vary in number, size, composition of neurotransmitters and hence can have *more or less, excitatory or inhibitory* effects and produce a weighted output from the pre-synaptic neuron to the post-synaptic neuron. Lastly the bias is biologically reflected by the activation threshold of postsynaptic neurons. If the resting membrane potential of neurons is around -70 mV (varies across neuron types) and the activation threshold around -55 mV (where an action potential is released), then the bias could be -15 mV where the activation threshold is redefined at 0 mV . Hebbian learning [99] serves as the biologically plausible model for the learning rule in the early Perceptron, where long-term potentiation would strengthen and long-term depression weaken the influence of the pre-synaptic neuron onto the post-synaptic neuron (reflected in the update rule see Footnote 5).

One of the most prominent issues in using the Perceptron as an analogon for human brain function is that actual biological neurons expose a much more complex response pattern (the action potential) compared to the single unit on/off response produced by the Perceptron. Action potentials alone vary significantly in their temporal dynamics, depending on the cell type, composition of membrane proteins, neurotransmitters, sub-threshold currents or ion concentration in the extracellular space [100]. Furthermore, mammalian neurons emit ac-

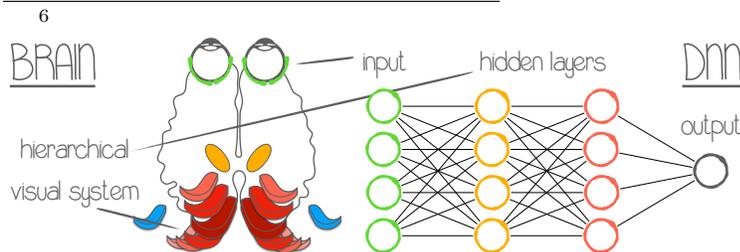


tion potentials spontaneously, even in vitro [101]. Additionally, the Perceptron does not incorporate feedback or recurrent connections, learning is exclusively driven by the external input and despite large amounts of training data, it is only capable of linearly separating two classes of (simple) input patterns.

Shortcomings in the simulation of the temporal dynamics have been addressed using spiking neurons that form a spiking neural network (SNN). Simple leaky integrate and fire (LIF) neurons [102] integrate multiple inputs and make use of a decay function that lowers the (simulated) membrane potential over time in order to mimic simple neuronal temporal dynamics. One of the most sophisticated (and computationally expensive to simulate) models - the Hodgkin–Huxley model [103] - uses a set of four biologically plausible coupled ordinary differential equations and is able to closely resemble biologically plausible action potentials and population response patterns. While the LIF neuron is highly efficient in terms of computational cost, it does not produce biologically plausible temporal dynamics on the single neuron level. A third model - the Izhikevich model [104] - is highly efficient in terms of computational cost and produces biologically plausible results. However, the formulation of the model itself does not find any correspondence in biological processes at all. Irrespective of the underlying neuron model however, it has been demonstrated for a large variety of applications that SNNs are capable of modelling the response dynamics of biological neural networks to a high degree of accuracy. Using a simplified version of the Hodgkin–Huxley model [105], Beyeler et al. (2014) [106] simulated the spike time temporal dynamics of motion selectivity in V1 and the middle temporal (MT) visual area. Their model successfully reproduced behavioral responses, exposed pattern and direction selectivity, as well as biological plausible firing rate tuning curves for speed selectivity. Furthermore, biologically plausible SNNs have been successfully used to mimic visual attention [107], spike-time dependent plasticity [108], learning in general [109], speech recognition [110] and robotic control [111]. For modelling large scale networks - such as the human brain - the dynamics of entire neuronal populations are modelled using mean field or neural mass models. The fundamental idea behind those models is the assumption that on a macroscopic scale the population dynamics of SNNs can be approximated to a high degree of accuracy by a somewhat overall network dynamic. Whole brain simulators, such as "The Virtual Brain" (TVB) [112] employ neural mass models to simulate different modalities of neuro-imaging data, such as fMRI, EEG or MEG data (see also "Functional and effective connectivity", WHAT IS ... fMRI on page 8, WHAT IS ... EEG on page 20 and WHAT IS ... MEG on page 22).

Not only does the Perceptron expose very low biological plausibility (no spiking, no recurrent or feedback connections, etc) its classification performance could be considered rather poor. While biological plausibility remains low, at least the poor performance of the Perceptron could be overcome by effectively stacking multiple neuronal network layers (mul-

multiple Perceptrons) hierarchically on top of each other ⁶. The resulting DNN is capable of modelling a much larger feature space. Additional layers between input and output neurons are called hidden layers, because they are not accessible to the observer and the term *deep* refers to the increased number of hidden layers. Furthermore, the way DNNs are trained has been immensely improved using techniques, such as stochastic gradient descent [113] or error back-propagation [84] and the introduction of new types of layer models, such as convolutional layers [114] or long short-term memory (LSTM) units [115]. Modern DNNs are able to not only meeting human level performance in certain tasks, but are furthermore capable of exceeding it. Famously, AlphaGo Zero and its predecessor AlphaGo achieved superhuman performance in playing the game of Go more than five years ago [116], a task deemed impossible to achieve for the foreseeable future. Although AlphaGo is made up of three components of which only one comprises a convolutional neural network (CNN), it has been this particular improvement to classical Go playing AI systems that allowed AlphaGo to achieve superhuman performance [117]. Especially computer vision systems highly benefited from improved DNN architectures over the past decade. In 2009, the ImageNet database - until then the largest database of labelled images - was released, containing 1000 different categories of images [118]. With this database, the authors challenged artificial intelligence researchers to compete over classification performance. Success was measured in the top 5% error rate, which reflects how often the target image category was not among the top five predicted results. In 2010, this error rate was at about 28% using the NEC-UIUC network [119], but soon after dropped to 16% with the introduction of AlexNet [120] in 2012, to 7% after the release of GoogleNet [121] in 2014 and finally to below 4% using the ResNet architecture [122] in 2015, which is considered better than human level performance [123]. Especially CNN architectures proved well performing for image recognition tasks. Naturally the question arose if DNNs and in particular CNNs resemble some kind of functional processes that are similarly implemented in (human) brains. After all, the visual system of mammals could be coarsely expressed in terms of a DNN architecture: the retina would resemble a linearly connected input layer and the hierarchical organization of the visual system the consecutive depth in the neural network. Seeliger et al. (2021) [124] trained a CNN, set up to partly mimic the human visual stream - retina, thalamus, V1, V2, V3, MT and fusiform face area (FFA) - to map the (original) stimulus input to the human BOLD signal. A single subject, watched 23.3 h of movie while fMRI data was recorded.



Stacking multiple neural network layers on top of each other before computing the output, creates a *deep* neural network and massively increases the computable feature space. Note, that this is coarsely similar to the hierarchical organization of the human visual cortex.

The CNN was trained to reproduce BOLD data from a re-scaled version of the very same visual input the subject had perceived. Afterwards, channel weights of the convolutional layer resembling V1 were extracted and visualized. Among a set of unexplainable channel visualizations, the authors found a set of stripy patterns that varied in spatial frequency and orientation, similar to patterns V1 preferably responds to in mammals (see Structural aspects of the brain). Many of the patterns furthermore exposed a temporal component in feature processing. In another study, a pre-trained CNN (VGG-S [125]) was compared to human MEG data obtained using a visual object recognition task [126]. It was found that the stimulus representations from the CNN could predict the MEG signal spatially and temporally. Thereby, early CNN layers better predicted MEG data from visual area V1 early after stimulus onset, whereas stimulus representations in deeper layers of the CNN predicted better the MEG signal from higher order visual regions later after stimulus onset. The idea of a computational familiarity between brain processes and activity patterns in deep convolutional neural networks, receives further support from Eickenberg et al. (2017) [127]. Again, fMRI data from visual regions low in the visual hierarchy could be modeled best, by early layers of a CNN, whereas higher order visual regions received higher prediction scores from activity patterns in deeper layers of the CNN. Additionally, a clear separation between dorsal and ventral visual stream, as well as high hemispheric symmetry could be observed, which was interpreted by the authors as a sanity check against spurious results. However recently, the convergence of scientific opinion towards DNNs reflecting an appropriate model for human vision has been challenged. Sexton and Love (2022) [128] report that the resemblance of cortical and DNN architecture and activation might reflect a methodological artifact. They used data of two fMRI studies in humans and one study involving macaque monkeys. But rather than applying aforementioned techniques to measure correspondence that focus on overall explained (or shared) variance, the authors applied an activity measure that only included task-relevant activity, which made typical brain-to-DNN correspondence patterns vanish in favor of deep layers that explain the data in any brain region better than other layers. Not only seem different relational measures produce strikingly different results, but furthermore the key driving factor for resemblance has been found to be network dimensionality [129]. This finding is remarkable, because it challenges the idea that DNNs and brains process (i.e. compress) the feature space similarly. However, if the explanatory power increases by increasing the feature space of the DNN then this view can be seen as highly questionable. If feature compression is compared between ANNs and the brain, but feature space inflation actually increases correspondence, then the features to be compared might actually be different from feature compression. A similar point is brought forward by Schaeffer et al. (2022) [130]. In a nutshell, the authors argue that the loss function that is computed in the brain cannot be inferred from the data, since multiple loss functions could have the same local minima in the resemblance function, making post-hoc parameter choices being overly influential on simulated brain activity. Nonetheless, all of the above mentioned critical papers suggest that those resemblance problems could be overcome by either finding

appropriate statistical measures and or by a rigorous implementation of biological plausible network architectures and parameter choices.

While comparing activation patterns of trained ANNs and brain data provides potential insights into the content of cortical processing, principles of how information is organized or exploited can be compared as well. As described above, error back-propagation and recurrent connections are often missing in DNN models. However, the brain is connected bi-directional and feedback related signalling explains a substantial amount of even early sensory neuronal data [69]. According to the predictive coding framework [131] (see WHAT IS . . . on page 33 for a brief overview about core principles of this framework), the brain creates predictions about upcoming states of sensory processing which are matched against sensory evidence. The amount of mismatch is reflected in the prediction error which in turn is used to refine later predictions. Choksi et al. (2021) [132] added this model to standard feed-forward image classification networks (such as VGG16). Each convolutional layer would be connected to a generative layer, which aims to predict the respective input to the convolutional layer. The mismatch between prediction and evidence would provide the loss function for the respective generative layer, whereas remaining layers are trained regularly. During each training iteration, the generative layers were iteratively updated until convergence. The main benefit of the augmented model has been the increased robustness against adversarial attacks. Those attacks target specific weaknesses of the network and can e.g. be made of very low value (not human detectable) but high impact noise that is added to the image data. Not only implies this finding a potential gain for AI research, but furthermore provides a plausible argument (not evidence) for a similar mechanism being at play in the human brain.

Both, SNNs and DNNs provide valuable models for task related activity in the visual cortex. Nonetheless, as Schaeffer et al. (2022) [130] phrased it, there is "no free lunch from deep learning in neuroscience". The brain however is a neural network and both, neuroscience and research on artificial neural network could derive mutual benefit by bringing both scientific disciplines closer together. Even fundamentally difficult problems like the creation of general AI could be approachable using neuroscience inspired ANN models [133]. The Neocognitron [134] for instance was directly inspired by the observations of Hubel and Wiesel on cortical response patterns of cat visual cortex, when exposed to differently oriented bars [30]. Simple feature selective cells (S-cells) and complex cells (C-cells) responding to specific activation patterns of S-cells, are stacked to combine simple features (e.g. bars) into complex features (e.g. corners). Based on the Neocognitron, CNNs could be seen as a direct consequence of insights about the visual cortex [135]. On the other hand do ANNs not only provide models and testable hypotheses, but furthermore allow for the development of new analysis strategies that enable rapid progression in brain computer interface (BCI) research [136–138], improvements of functional prostheses [139, 140] and neuro-imaging in general [141, 142]. Here however, neural network models are used and described solely in the context

of cortex models and applied AI research based on DNNs will be neglected. One major challenge when employing ANNs as brain network models, is the implementation of temporal dynamics and connectivity between neurons or populations. As the introduction example of the *Hydra* showed, even a fully known (spatially) model, cannot be understood without an understanding of the temporal and functional dynamics of the neuronal communication, at least on the level of canonical microcircuits.

What is ... electroencephalography (EEG)?

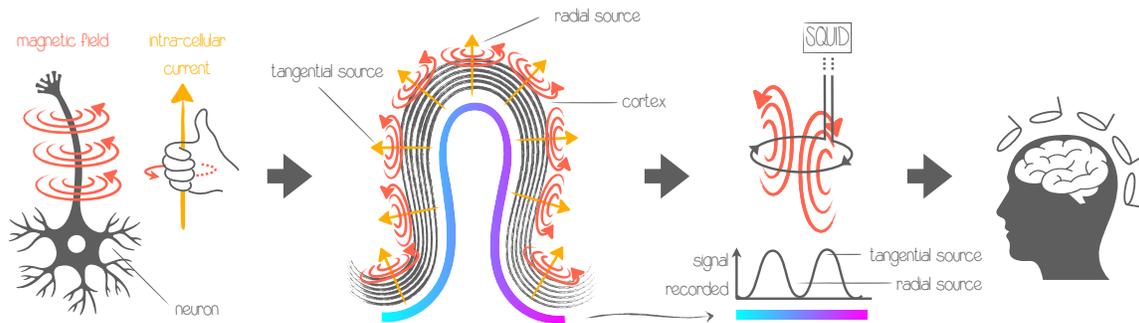
Electroencephalography (EEG) is used to measure small currents, generated by the brain, on the scalp. Synaptic signalling is tightly linked to the exchange of molecules and ions at the synaptic cleft (see WHAT IS ... on page 5). An excitatory post-synaptic potential (EPSP) causes the influx of positive electric charge, leaving the remaining extracellular space charged relatively negative. On the other end of the neuron (i.e. axon) a surplus of positive charges in the extracellular space - relative to the first region - creates an electro-chemical gradient in the extracellular space. This gradient causes an extracellular current flow towards the area of the initial influx. In case of an inhibitory post-synaptic potential (IPSP) the direction of current flow is reversed. The movement of electrical charges spreads through the tissue which is known as volume conduction. Thereby, different tissues (e.g. brain, skull or scalp) modify the current flow based on local changes in conductivity. Hence, as compared to MEG, EEG relies on sophisticated tissue models for source reconstruction. The forward current flow from the neuronal source to the electrode attached on the scalp is modelled and the inverse is estimated by maximize the similarity between the forward generated data and the recorded data, given a specific source location and orientation. EPSPs thereby are theoretically deflected positively, whereas IPSPs are deflected negatively in the scalp EEG. In reality, the actual underlying neuronal deflections are hard to pinpoint from scalp EEG recordings, due to the vast number of neurons averaged in the signal. This info-box was first referenced on page 16.

1.3 Temporal aspects of the brain

Cortical neuronal communication comprises at least two aspects: local or proximal connections and more widespread (e.g. inter-area) connections. So far, only structural aspects, such as canonical microcircuits (see "Canonical microcircuits: cortical compute modules") have been discussed, mainly on the basis of anatomical connections. Temporal dynamics however form *the* fundamental functional components that enable information encoding in the first place. Neuronal communication is based on chemical reactions between synapses. Once a certain electro-chemical threshold is hit, a temporally precisely tuned electro-chemical signal cascade - the action potential - is initiated, which in turn causes a chemical reaction in the synaptic junction (see WHAT IS . . . Neuronal activity on page 5). Within the realm of this work however, bio-chemical processes will widely be ignored in favor of information theoretical properties. The term neuronal communication will hence always refer to an exchange of information between neurons (in the form of spikes). Temporal dynamic aspects of neuronal communication on the level of single neurons as well as individual neuro-neuronal connections will be reduced to information theoretical or computational aspects too. Neuronal activity, reflected in the BOLD signal or electromagnetic fluctuations of intra- and extracellular current flow (LFP, EEG, MEG, etc.) [143], is generated by the joint activation of many thousands of neurons [29]. Naturally, computational models at this scale cannot include individual neuronal dynamics. Furthermore, fine grained temporal dynamics that simultaneously happen across many thousands of neurons inevitably indicate highly non-random coordinated behavior. Phases of excitation and inhibition between neurons must be coordinated, such that neither important signal patterns get canceled out (e.g. by accidentally incoming inhibitory activation), nor an over-activation of wide-spreading activity (as observed in epileptic seizures [144]) disturbs the balance of the entire system.

In terms of Shannon entropy [145], information between neurons can only be transferred if for a period of time, the probability of any given neuron to fire is neither 0, nor 1. Both would either yield mathematical implausible results or do not convey any information. This insight implies already - to some extent - varying firing patterns for specific pieces of information that are transferred. Furthermore, since the input to the brain (sensomotoric, visual, olfactory, etc.) is physically and perceptually structured, a correlation between signals of e.g. neighboring retinal cells would be expected. In addition to non-random temporal coding and spatially correlated neuronal signals, the temporal structure of the environment needs to be accounted as well, which determines a temporal correlation of sensory input signals too. Spatially correlated non-random spiking activity synchronizes large parts of the respective neuronal population. Empirical data shows a large variety of rhythmical patterns for different frequency bands that could be successfully linked to task related and resting state brain functions in mammals [146]. This finding however is not limited to large complex networks, since even the neural net of the *Hydra* with only a few hundreds or thousands of

What is . . . magnetoencephalography (MEG)?



Magnetoencephalography (MEG) is a neuro-imaging technique that allows the recording of magnetic fields produced by neuronal activity (see WHAT IS . . . on page 5). Orthogonal (circumferential) to the direction of the intra-cellular current flow within an active neuron a magnetic field is created (the direction can be inferred using the right hand rule). Since neuronal populations in the cortex are mostly aligned in columns, the direction of the current flow for large populations of neurons is similar and hence the magnetic field of similar direction. The fields of thousands of neurons can be detected using superconducting quantum interference devices (SQUIDS). Thereby, the magnetic field (red) induces an electric current in the pickup coil (black circular arrow). Respective field strengths detected vary in the fT (femto-Tesla: 10^{-15}) range. Due to the way the signal is measured, neuronal sources that are exactly orthogonal to the sensor (radial sources) cannot be detected. Using hundreds of sensors, neuronal generators can be inferred by e.g. modelling neuronal sources as dipoles and estimating which dipole at which location inside the brain would cause which respective sensor level data. The model (source location and orientation) with the highest explanatory power is assumed to be the main contributor to the data. See also WHAT IS . . . EEG on page 20 This info-box was first referenced on page 16.

neurons exposes slow rhythmic fluctuations [147]. It has been hypothesized that neuronal oscillations couple neuronal activity, such that phases of low and high excitability coordinate cortical information flow [148] and recruit neuronal populations to form widespread computational networks [91]. Hence, current theoretical models on neuronal communication are often tightly linked to functionally relevant oscillatory patterns. Furthermore, cortical oscillations are thought to play an important role for local computations on the level of canonical microcircuits [82]. Even though the temporal dynamics of neuronal signalling are within the ms range, temporal dynamics reflected by the BOLD signal in the s range, still provides insight into cortico-cortical connectivity.

Functional and effective connectivity

Fundamentally, functional and effective connectivity reflect the idea that (some) cortical functions are spatially highly distributed across cortical neuronal populations. For example, it has been suggested that selective attention is implemented using a cortical network that spans fronto-parietal regions [149]. Rather than implementing every cortical function in a separate module, multiple cortical modules interact, such that emergent behavior can occur. Functional and effective connectivity reflect the attempt to infer functional networks of spatially separated neuronal populations from time resolved neuro-imaging data. Whereas functional connectivity reflects a mere temporal correlations of jointly activated regions, effective connectivity considers causal influences (e.g. determined by anatomical connections) as well [150]. Traditionally, functional connectivity has been assessed in form of correlograms reflecting the cross-correlation between two region's spike trains, but was soon extended to temporal correlations between brain regions using whole brain neuro-imaging methods, such as positron emission tomography (PET) [151]. Modern estimates of the functional relationship between activity patterns across spatially separated brain areas can be divided into directional and non-directional methods, as well as model based and model free methods [152]. *Directional* measures, such as Granger causality [153] assume (in the given context) that if one brain region drives another brain region, the time series signal should be similarly reflected in both regions, but should occur earlier in the driving region. The lag is explained as the time the neuronal signal travels physically through the brain. Hence, Granger causality derives directionality from the sign of the lag between signals. If region A drives region B, then a similar time series could be observed in region B as compared to region A, but shifted forward in time. However, if region B would drive region A, a similar relationship between the signals might be observed as well, however the signal of region B would be shifted backwards in time compared to region A. More broadly, Granger causality reflects how well one time series is able to predict another time series in the future. Causality thereby is assumed, if the time lag between predicting and predicted time series can be aligned with plausible hypothesis about the underlying mechanisms. For the interaction between multiple brain areas, a lag in the range of milliseconds might yield higher plausibility than a lag time in the order of minutes. In turn a time lag of close to 0 *ms* might as well be seen as biological implausible and hence true causality would be less plausible. In turn *non-directional* measures, such as mutual information [145], ask how much information about one data set can be obtained by observing another data set. One major advantage of mutual information compared to the Pearson correlation coefficient is that the former is sensitive to statistical dependencies that are not reflected in the covariance between the data sets as well [154], whereas the latter depends on exactly this measure. However, as the name already suggests, the shared information content is mutual and thus conclusions about directionality cannot be inferred. *Model based* as compared to *model free* measures of functional connectivity, assume a specific known relationship (that is the model) between

cause and effect. For instance the Pearson correlation coefficient assumes a linear relationship between multiple data sets (model based), whereas mutual information does not (model free).

A relatively new approach to estimating functional connectivity in terms of effective connectivity is using a dynamic causal model (DCM) [155]. Over the past two decades, this approach has been widely popularized, due to its capability to capture non-linearities in the data and the possibility to link observed data to biologically plausible constraints and the measuring process itself [152]. Dynamic causal modelling can be viewed as a highly deterministic approach, where every current state could be inferred from previous states. Thereby dynamic activation states are causally affected non-linearly by perturbations to the system (e.g. by sensory input). From those perturbations, model parameters for the coupling of e.g. two brain areas can be inferred. In its core the DCM approach is hence a generative model, where the error between model predictions and measured data is minimized in order to estimate respective model parameters. While the original paper by Friston and colleagues (2003) [155] mostly addresses fMRI data, DCM can be used to explain highly temporal resolved EEG and MEG data as well [156, 157]. By incorporating neural mass models including feed-forward, feedback and lateral connections (see "Canonical microcircuits: cortical compute modules"), David et al. (2006) [157] demonstrated that DCM parameters on cortico-cortical connectivity can be inferred using Bayesian inference. By incorporating neuro-biological plausible priors into the model, event-related potentials (ERP) and event-related field (ERF) as macroscopic brain responses to specific tasks could be modeled and underlying connectivity patterns inferred. However, not only allows the DCM approach to infer the connectivity patterns that gave rise to certain macroscopic observations, but furthermore to infer connectivity changes from the respective stimulation by the task [156]. While fMRI data inherently constrains the spatial parameter space by tying exact locations within the brain to the (slow) temporal dynamics of the BOLD signal, applying DCMs based on EEG or MEG data requires detailed assumptions about the respective anatomy and respective estimated sources of neuronal activity measured at EEG or MEG sensors using e.g. source reconstruction techniques. Since an infinite configuration of neuronal source models could potentially give rise to the observed data at the sensor level for EEG or MEG data, parameter estimation inference using DCM, can in some cases yield ambiguous results [158]. This means that it is not guaranteed that the estimated parameters indeed reflect underlying processes, which requires additional constraints (e.g. by the research hypothesis) to support results. Functional and effective connectivity - as a model for long range neuronal interaction - and frameworks, such as canonical microcircuits (see Canonical microcircuits: cortical compute modules) to describe local interactions, could serve as the basis for simulation studies putting respective frameworks to the test. Neuroscientific evidence and theoretical models could be cast into ANN simulations, which would not only allow to test respective model predictions, but furthermore would potentially enrich the pool of testable hypotheses for in vivo experiments. Especially neuronal oscillations thereby might play an important role.

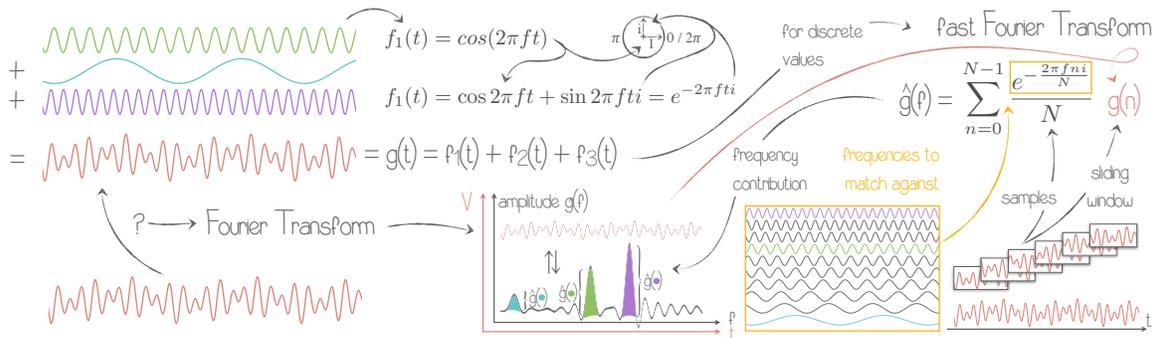
Neuronal oscillations

The term neuronal oscillations refers to rhythmic activation patterns of neuronal populations that have been observed in many species [146]. In order to investigate neuronal oscillations in humans, multiple invasive and non-invasive techniques that expose a high time domain resolution, such as electroencephalography (EEG), magnetoencephalography (MEG), electrocorticography (ECoG), local field potentials (LFP) and others have traditionally been used. The raw time domain electromagnetic signal is recorded using specific sensors and afterwards decomposed into its spectral components using e.g. a Fourier analysis (see WHAT IS ... time-frequency decomposition on page 26 for an overview of how a signal can be decomposed into its spectral components using a (fast) Fourier transform).

Neuronal oscillations around 18 Hz have been linked to functional processes in the olfactory system of *Honey Bees* [159]. This remarkable finding indicates a very general role of neuronal oscillations going way beyond their role in complex information processing as observed in mammal brains [146]. On an even smaller scale, α oscillations (around 10 Hz) have been linked to differential responses to a variety of odors in *Drosophila* [160]. Whether neuronal oscillations are an inherent feature of *every* (spiking) neural network remains to be understood, however simulation studies indicate that spiking neural network models are capable of producing fast oscillations that are dependent on the spike timing of the neurons and slow oscillations that are dependent on the membrane time constant [161]. Depending on the frequency band (and sometimes based on functional differences or the cortical location to which they are attributed), neuronal oscillations are categorized in separate bands: slow waves (below 1 Hz), δ (1 – 4 Hz), θ (4 – 8 Hz), α (8 – 14 Hz), β (14 – 30 Hz), γ (30– \approx 120 Hz) and fast oscillations ($>\approx$ 120 Hz). While slow waves and δ oscillations have been mostly linked to sleep, recent work suggests a potential mediating role of δ oscillations in sensory processing. Specifically, it has been suggested that δ encodes contextual sensory information and thereby constrains the sampling of this information with respect to the temporal domain [162]. Since δ oscillations have been only of little interest for studying awake perception, the literature linking δ and (awake) cognition is rather scarce. In turn θ , α , β and γ oscillations have been linked to a variety of cognitive, perceptual or motor related processes for many decades and will thus be discussed in more detail below.

Note however that the term neuronal *oscillation* is a bit misleading. Even though it does not imperatively imply a sustained oscillation, it nevertheless has been interpreted for a long time more or less exclusively this way. Especially for high oscillations (e.g. β or γ), it is believed today that many seemingly sustained oscillatory response patterns are actually the result of bursts () of neuronal activation that are averaged across multiple trials, inducing the illusion of a more or less sustained oscillatory response [163]. Furthermore, thalamic burst activity (specifically in the LGN) inducing slow frequency oscillations

What is ... time-frequency decomposition?



To analyze the frequency content of a neuronal signal, one of the major strategies is the use of a time-frequency decomposition with the help of a fast Fourier transform (FFT). The basic idea is that any signal can be described by the weighted sum of a number of fundamental sinusoidal frequencies, with $g(t) = a \cos 2\pi ft + p$ where a is the amplitude, f the frequency, t the time and p the phase of the oscillation. For $a = 1$ and $p = 0$, the oscillation can be described by f revolutions around the unit circle ($r = 1$) in t time. Every point p_t of $g(t)$ hence can be described by two dimensional coordinate that lies on the unit circle. Since $r = 1$ the movement around the unit circle can be described by a complex number (in complex plane), where the real part corresponds to the "x coordinate" and the imaginary part to the "y coordinate" that describes the points. This is known as Euler's formula: $e^{-2\pi ft i} = \cos 2\pi ft + \sin 2\pi ft i$. If an unknown signal $g(t)$, composed of multiple base frequencies is multiplied by a function with a given base frequencies $e^{-2\pi ft i}$, then, the resulting integral (the sum under the curve) reflects the contribution of that frequency f . The Fourier transform becomes:

$$g(\hat{f}) = \int_{-\infty}^{\infty} g(t)e^{-2\pi ft i} dt \iff g(t) = \int_{-\infty}^{\infty} \hat{g}(f)e^{2\pi ft i} df$$

For finite length signals the discrete or fast Fourier transform (FFT) $\hat{g}(f) = \sum_{n=0}^{N-1} e^{-\frac{2\pi f n i}{N}} / N$ is used to obtain the average base frequency contribution for a respective signal with N samples. To furthermore obtain *time*-frequency resolved data, a sliding time window of a certain length and sliding step size is used. For each sliding step, the FFT is computed separately. Since the maximum frequency that can be resolved is half the sampling rate (Nyquist frequency), and the lowest frequency is bound by the sliding window width, time and frequency resolution are inversely related for this approach. Shorter windows allow for a better time resolution, whereas longer windows increase frequency resolution (and sensitivity to low frequencies). This info-box was first referenced on page 25.

(i.e. α and θ) has been considered a main driver for EEG level α and θ oscillations [164]. Here, neuronal oscillations are often displayed as quasi-sustained for illustrative purposes. Nevertheless, proposed mechanisms for neuronal communication based on neuronal oscillations (see sub-sections below) do not imperatively imply nor require a sustained oscillatory activity. Due to their importance especially for visual processing in the human neo-cortex, general functional roles of α , γ , θ and β oscillations will be highlighted in this section, before they are respectively linked to theoretical computational models.

Alpha The first oscillatory pattern that has been discovered using EEG was a ≈ 10 Hz rhythm that became prominently visible over occipital sensor sites when participants closed their eyes [165]. Hence, the name α oscillation. It has been argued that α oscillations reflect an inhibitory cortical "idling" state that is suppressed if some kind of cognitive "load" is applied [166]. A reduction of α power after the onset of an event was explained by asynchronous firing patterns (as compared to idling) of the respective neurons which process the stimulus and was therefore called event-related de-synchronization (ERD). Enhanced or even right on enabled information processing by de-correlating neuronal firing can be explained by means of information theory. In information theory, one classical measure to express the amount of information within a system is called Shannon entropy [145]. Given a certain state space (e.g. potential neuronal firing patterns in a network), Shannon entropy describes the amount of information that can on average be conveyed by an observation of one of those states (i.e. an event). More "surprising" events - events that are less likely - thereby convey more information than less surprising events. In other words, a predictable outcome is little informative. For example, when playing the game "who am I?", a yes answer to the question "Am I a human?" is less informative than a yes answer to the question "Am I a female human?". Without any prior knowledge, the second question reveals more information if answered positively, however a positive outcome is less likely. In contrast, an event that could be predicted with absolute metaphysical certainty would yield no information at all (e.g. "Can my identity be revealed by yes or no answers to questions within the context of 'who am I'?"). Applied to neuronal oscillations, a neuronal state that could be predicted by the previous state would yield less information than a state that is less predictable. Spatially or temporally correlated signals of a population of neurons would hence contain less information than a de-correlated pattern. A purely sinusoidal α oscillation across all neurons within a population, phase locked to each other, would hence contain (almost) zero information. From the perspective of a neuronal population, an unexpected event (e.g. the onset of a stimulus) would be reflected in increased entropy and the de-correlation of the signal follows consequently. The change of e.g. retinal input, when participants open their eyes in the classical experiment by Berger [165] requires the encoding of new information, which necessarily must be reflected in an increase in entropy. Thus, the brain would leave its

cortical "idling state" and an ERD could be observed. This very principle has been linked to memory formation and retrieval as well, where decreased α (and β) power was associated with more successful memory formation, explained by the increased amount of information that could be extracted due to the ERD in neuronal populations [167]. If α band ERD reflects increased information processing, then α band event-related synchronization (ERS) might as well reflect the converse. In a complex environment, successful information suppression can contribute to the optimization of the overall information processing pipeline, since otherwise cognitive resources would be allocated to potentially meaningless or irrelevant stimuli. The interplay between α ERD and ERS - enhancing the processing of relevant and suppressing the processing of irrelevant information - has thus been considered an important feature of cortical information processing reflected in the α band [48, 168–170]. Indeed, it could be demonstrated that occipital α band power positively correlates with the capability of successfully suppressing irrelevant information in visual tasks [171, 172]. Furthermore, those suppression effects can even be observed for lateralized stimulus presentation, such that α power increases over occipital regions ipsi-lateral to the visually presented target stimulus and contra-lateral to visually presented distractor stimuli [173]. Moreover, Sauseng et al. (2009) [174] showed that α ERD and ERS effects can be artificially enhanced using repetitive transcranial magnetic stimulation (rTMS). In a lateral attention, short term memory task, participants were asked to centrally fixate a screen, displaying one array of colored squares on either side. A previously presented cue indicated which of the two arrays should be retained in memory, whereas the array on the other side should be ignored. Afterwards, participants were asked to compare the memorized array to a subsequent probe. Using rTMS, electric pulses (or sham stimulation) in the α range have been applied to participants posterior parietal cortices. It was found that rTMS stimulation during the retention interval in the α band (enhancing ERS) over ipsi-lateral posterior parietal cortices relative to the side of attention significantly increase the participants performance, even over sham and vertex rTMS stimulation. In other words, visual short term memory performance was enhanced in that condition. Conversely, contra-lateral rTMS stimulation significantly decreased performance. Those findings were interpreted, such that ERS enhancement via rTMS over parietal regions ipsi-lateral to the attended side led to better inhibition of task irrelevant information, resulting in an increase of task performance, whereas a disturbance of ERD via rTMS over parietal regions contra-lateral to the attended side, disrupted successful stimulus processing (or encoding). In the light of Shannon entropy, an artificial synchronization of neuronal populations processing task relevant information limited - to a certain degree - the amount of information that could be extracted, by lowering neuronal entropy. The nature of α oscillations can hence be described as rhythmic fluctuations of cortical inhibition, coordinating neuronal processing by modulating post-synaptic excitability. Recent findings suggest that α band oscillations do not reflect an active mechanism to implement selective attention and that findings on rhythmically activity in the α band might be instead a consequence of selective attention [175]. Additionally, stimulus enhancement rather than

distractor suppression could explain classical findings just as well [176]. Moreover, it could be demonstrated that distractor suppression is shaped by previous occurrences of respective distracting information on spatial and feature level [177]. Again, a simple suppression of distractor stimuli could not explain those findings. Instead it has been suggested that α band oscillations hamper distractor processing by stabilizing sensory processes in a particular configuration [178] which could e.g. serve as cause for bi-stable visual perception phenomena [179]. This would be in line with the idea of α oscillations implementing spatio-attentional sampling [180]. Each excitable phase of the α cycle would vary in length, depending on the exact frequency. Hence, a modulation of α frequency could in this context be interpreted as the duration a communication channel is set to "open". Stimuli relying on a temporal integration would hence provoke slower α band activity that in turn would be different in frequency as compared to the separate distractor signals [181]. Distractor suppression would hence not be implemented as an active process, but would rather be seen as a consequence of the non-recruitment of neuronal populations preferring task irrelevant features. See also "Gating by inhibition" for a discussion about the inhibitory nature of α oscillations, "Communication via nested oscillations" where the role of α oscillations is discussed within a joint framework developed by Bonnefond et al. (2017) [91] and "Feature specific neuronal oscillations in cortical layers (in prep)" where results of the present project - with respect to the differential role of α - are rolled out.

Gamma Ranging from $\approx 30\text{ Hz}$ to $\approx 120\text{ Hz}$, γ oscillations have been linked to stimulus related activity and was thought to reflect feature specific activation [182]. It has been hypothesized that simultaneously activated feature specific neurons (e.g. neurons responding to a specific orientation [30]) must be temporally aligned to integrate multiple information streams. Rhythmical activity in the γ band was attributed to the process of feature binding to create a coherent percept of visual input [183, 184]. Time-frequency resolved EEG data exposed a significant peak in the power spectrum of frequencies between 30 Hz and 60 Hz , when participants were presented with the Kanizsa illusion (). No such power increase has been observed for a version with rotated pacmans (no illusion effect). The authors conclude that fast γ band oscillations reflect the formation of coherent task-relevant object representations. This view was extended to the idea that γ band activity reflects a general computational mechanism that coordinates spike timing to bind widespread cortical activity into functional networks that integrate distributed responses [185]. However, later evidence suggests that not binding by synchrony (causing γ band oscillations) but rather binding by firing rate (firing not necessarily synchronous but statistically equally likely) could explain feature binding or grouping [186]. Oscillations in the γ band have instead been linked to more local processes that are translated into firing rate changes later in the processing pipeline [187]. Other theories view γ band oscillations as markers for stimulus feature related feed-forward processing [91] or question the nature of γ as a "true" oscillation in general, by

suggesting γ might not play a role at all in cortical processing but nevertheless can serve as a marker for local neuronal interactions [188]. Findings using ECoG data from humans revealed a stimulus dependent γ band response in primary visual areas V1 and V2 [189]. Properties of the stimulus, such as spatial frequency or color reliably modulated the γ band response. For noisy stimuli, the response has been reported to be much more broadband and lower in peak amplitude as compared to grating stimuli. Complex objects in turn elicited both, a small broadband or a small broadband and a large narrow band γ response depending on the exact stimulus. Since noisy and complex images are reflected in a more broadband power spectrum likewise, broadband γ could indicate stimulus complexity to a certain degree. Functionally different roles for broad and narrow band γ band oscillations have also been reported by Bartoli et al. (2019) [190]. Narrow band γ was found to be specifically tuned to specific stimulus features (structure and color), whereas broadband γ could not be linked to specific stimulus features in the same way. It has been hypothesized that broadband γ could serve as a proxy for neuronal spiking, due to its high similarity to multi unit activity (MUA). It has further been argued that high frequency broadband γ oscillations might reflect a bias in the data. According to this explanation, low frequency components of neuronal spiking patterns reflected in broadband γ , are actually shadowed by $1/f$ noise ⁷ in the power spectrum [192]. Hence, for lower frequencies, the signal-to-noise ratio (SNR) is too low for the broadband γ signal to be detectable which makes it only visible in high frequencies where the SNR is more favorable towards the signal. In mice, the strength of visual broadband γ LFPs is correlated with contrast, whereas narrow band γ LFPs are negatively correlated with contrast, but both components have been found to not be correlated with each other [193]. It has been argued that narrow band γ reflects feed-forward processing from the thalamus (specifically LGN) to V1, whereas broadband γ is associated with the activation of cortical inhibition networks. For a discussion of the role of γ band oscillations in cortical information processing with respect to inter-area communication and information transfer see section "Communication through coherence" (below). The exact role of γ band oscillations has been controversially discussed. However, common features that most hypothesis about the role of γ exhibit include feature specificity, feed-forward directionality of information flow and the link to selectively activated neuronal populations as a response to stimulation.

Theta Early theories on the function of θ oscillations (4 Hz to 8 Hz) have been mainly linked to hippocampal processes [194]. It has been suggested that θ band neuronal fluctua-

⁷The EEG or MEG power spectral density follows a $1/f$ shape. With increasing frequency f , the power spectral density $S(f)$ decreases with $S(f) = 1/f$ (for $f > 0$):  . Hence, the respective noise power spectral density is higher for low frequencies than for high frequencies as well. This behavior has not only be observed for neuronal membrane voltages and currents, but furthermore for rain fall, rate of traffic flow or currents in vacuum tubes [191].

tions in the hippocampus of rabbits are driven by "pacemaker" cells in the septum, which would provide "a sense of time" [195]. Thereby, θ would necessarily be independent of external events. This view has been challenged by observations that the θ phase changes as a function of behavior and location, linking θ band activity *particularly* to external events (spatial locations) [196]. It could be shown that indeed spatial location and not the temporal dynamics of behavior are causing a change in hippocampal θ band activity [197]. In addition, hippocampal recordings revealed a strong relationship between θ phase and the spatial location of freely moving rats, but again not with the temporal dynamics of behavior (e.g. motion) itself [198]. θ oscillations thereby were found to be closely related to the activity of location specific hippocampal neurons (place cells). Firing patterns for specific place fields⁸ were thought to be modulated by a compound wave () of multiple frequencies () and or phases () would determine the extent of the place field (size of encoding time window), whereas the average phase of the base oscillations would be reflected in the phase (and amplitude) of the compound oscillation, which influences spike timing. Thus, by shifting multiple frequency specific signals relative to each other in phase and frequency, spike timing is modified, such that information under the envelope frequency is grouped together. Information about the exact location would be uniquely coded by the joint phase of the compound oscillation, whereas the shift of the envelope frequency would determine directionality. The authors conclude that θ might reflect a general principle of information processing and storage in the hippocampus. On the other hand, θ was found to play a potential role in active sensing. It could be observed that the whisking frequency in rats [199] as well as saccadic eye movement in humans [200] are correlated with θ frequency. The general role of θ has since been attributed to information sampling. It has e.g. been argued that θ could play a main role in dyslexia. According to Goswami (2011) [201], dyslexia might be a result of poorly temporally organized data. Impaired θ phase locking in the auditory pathway was thought to lead to impaired separation of the auditory signal and thus a reduced ability to link auditory features to phonemes. Additionally, the temporal structure that θ provides in the hippocampus has been associated with episodic memory and *its* temporal structure [202], as well as with a general mechanism for encoding and retrieval. An increase in θ power (in terms of absolute power values and relative changes in form of ERS), was related to better memory performance [203]. Those findings have been replicated for long term memory tasks as well as for working memory tasks and were associated with multi-item organization or sampling [204]. The detection threshold of attended stimuli depends (among the α frequency) on θ oscillations, such that the stimulus detection threshold could be predicted by the EEG θ phase [205]. Using a fine grained behavioral experiment, Huang and Luo (2020)

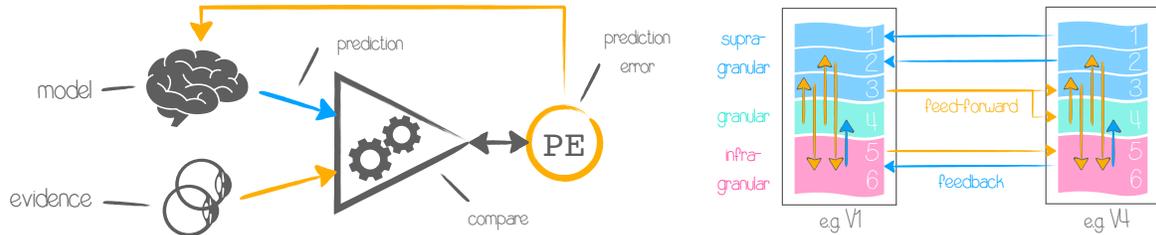
⁸Place cells are location specific neurons that first have been discovered in the hippocampus of rodents [197]. Each location specific neuronal population fires with higher probability to a given spatial location, known as the place field.

[206], demonstrated that high and low salience stimuli are processed at different relative time points with respect to stimulus onset. More specifically, behavioral stimulus detection rates fluctuated in the θ frequency range when represented as a function of the stimulus onset asynchrony (SOA). High and low salience stimuli are thereby processed at different phases, temporally prioritizing stimuli based on salience. Together, past and recent findings on the function of θ oscillations speak in favor of the idea that θ fluctuations reflect a temporally organized sampling of information uptake and encoding along its phase gradient. Gaillard et al. (2022) [180] argue, α nested in θ oscillations (see "Communication via nested oscillations" and "Coding via nested oscillations") reflects the shift of the attentional spotlight during continuous exploration of space by alternating between "exploration" and "exploitation" epochs. The possible separate role for θ in visual attention sampling [207] is complemented by the idea that attention θ acts alongside attention α [208, 209], were the latter is possibly nested in the former [210, 211].

Beta Between α and γ oscillations, the β frequency range spans roughly 14 Hz or 16 Hz to 30 Hz . β oscillations have classically been attributed to somatomotor functions [212] and origin [213] or motor imagery [214]. It has been observed that β power over central sensor sites (motor area) decreases shortly after the decision to execute a movement was made, reaching its minimum when the movement was executed and experienced a rebound effect after movement execution was finished [215]. The decision making process itself was linked to a fronto-medial β increase when inhibiting movement, which was reflected in a fronto-central β increase related to motor inhibition as well. Pre-movement related β power decreases (followed by post-movement rebound), could further be observed in anticipation of rhythmically occurring stimuli [216], implying a role of β in top-down related processes. Indeed, top-down related functions that involve rhythmic activity in the β band was found. Specifically, V4 and temporo-occipital (TEO) area feedback connections to V1/V2 in macaque monkeys have been related to differential response patterns, depending on the behavioral context (go or no go for specific stimuli) within the β band [217]. Those results have been argued to reflect feedback, based on behavioral anticipation of the predicted environment, modulating even primary sensory (visual) regions. Processes in the β band have hence be linked to "predictive" gain control [218] or feedback related activity within the predictive coding framework [82, 131] (see WHAT IS ... on page 33). Bressler and Richter (2015) [218] describe the role of β oscillations to represent feedback-directed context-dependent sensory expectations, originating in superficial layers of higher order regions that are connected to deep layers of lower order (or primary) sensory areas. This is in line with findings that investigated the link between predictive coding, where the prediction error was linked to γ band oscillations (feed-forward or bottom up operation) and the respective predictions to β band oscillations (feedback or top down operation) [219]. Oscillatory activity in the β band however, most likely is expressed in form of short lived burst activity of a specific waveform [220, 221]. Those

insights have been gained from single trial analyses, because after averaging over many trials a "sustained" oscillation is implied due to the different times at which β burst occur.

What is ... Predictive coding?



The predictive coding framework [131] treats the brain as a dynamic system where bottom-up sensory information and internally generated top-down predictions combined form a coherent perception. The mismatch between prediction and observation is called *prediction error*. Hence, the prediction error encodes new information. Both predictions and sensory information are constantly updated, which constantly changes prediction error values that in turn are used to refine predictions. The canonical microcircuits (see "Canonical microcircuits: cortical compute modules") model by Bastos et al. (2012) [82] - on the right side of the figure - assumes an interaction between multiple frequency bands (β and γ) to implement the predictive coding feed-forward and feedback network architecture. Feedback connections (blue arrows) would carry top-down predictions, that are compared against sensory evidence or the incoming signal from a lower cortical region to compute the prediction error. In turn the prediction error (orange arrows) is forwarded to higher order regions. Moreover, the predictive coding framework is an hypothesized implementation of Bayesian inference [222], The *Bayesian Brain* is thought to perceive the world through the lens of surprise, that is the mismatch between the predicted sensory signal and actual evidence. Predictions thereby are constantly adjusted in order to minimize the *surprise* (prediction error). Information thereby is represented in form of probabilistic distributions. Two main aspects have major influence on the prediction error: *precision* (rate of useful information out of all information extracted or *confidence*) and *recall* (rate of useful information that could successfully be extracted or *sensitivity*). Interestingly, some components could successfully be attributed to certain frequency bands: predictions: feedback directed β band activity [82]; prediction error: feed-forward γ band activity [82]; and precision: feedback directed α [223]. This info-box was first referenced on page 19.

The computational role of neuronal oscillations

Neuronal oscillations can be observed for a wide variety of tasks and in varying cortical regions and furthermore express a high variability within the frequency spectrum. Thereby, α and β oscillations are often linked to modulating processes operating in feedback direction, whereas γ oscillations have been linked to a feed-forward flow of information related to the stimulus itself. Since neuronal oscillations have been linked to a large variety of cortical functions, theories about general computational properties of rhythmic fluctuations have been gaining more and more attention. Here the most relevant to the presented project are rolled out.

Gating by inhibition

Cortical α oscillations (see also "Neuronal oscillations") have long been interpreted as "idling oscillation" [166]. This idea can be seen as a consequence of early findings by Berger (1929) [165], who observed a spontaneous synchronization of the EEG signal in the 10 Hz range over occipital areas, when participants closed their eyes. Observations that α power increased under high working memory load [225] and decreased when performing a visual attention task [226] support this idea, adding the notion of α synchrony reflecting an inhibitory "default state" that under cognitive load frees up resources. Those theories view cortical α oscillations as a global phenomenon [227] and - to some extent - as some kind of cortical baseline activity [228]. Around the same time, early theories of inter-area communication have emphasized that task related changes in cortical activity in one area, reflected by an increase in the BOLD signal in fMRI experiments, might be explained by task related connectivity changes between e.g. primary sensory areas and remote areas [229]. It has been hypothesized that the flow of information between areas is somewhat co-

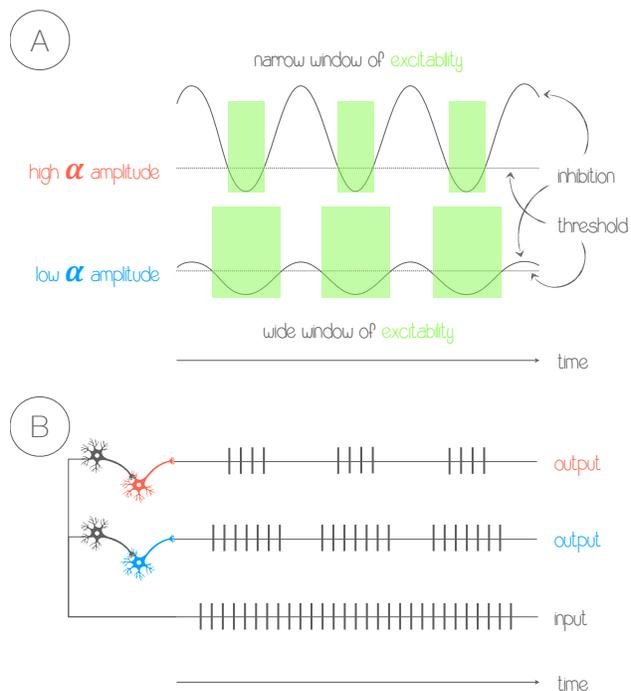


Figure 2: Gating by inhibition hypothesis. Figure adapted from Klimesch et al. (2007) [224]. **A) α amplitude dependent excitability.** The length of the window of excitability (green) depends on the amplitude of inhibitory α oscillations. The dotted line represents the inhibitory threshold at which incoming signals are blocked from traversing between areas. At high α amplitudes (red), a smaller fraction of time points fall into a phase where the inhibition is low enough to allow for the signal to be passed on. If the α amplitude is lower (blue), this window becomes wider and the signal can pass for a longer period of time during each cycle. **B) Inhibition dependent spike trains.** Given the respective excitable time windows derived in A), the signal of the gray neuron (bottom line of the spike trains), can be transmitted to a larger degree to neurons under low α inhibition (blue) as compared to neurons with high α inhibition (red).

ordinated to achieve *effective connectivity*. In conjunction, electrophysiological findings and fMRI data on cortical connectivity, led to the idea of cortical α oscillations reflecting gating processes to modulate information flow between areas [224].

The *gating by inhibition (GBI)* or *inhibition timing* hypothesis states that α oscillations are of inhibitory nature allowing, information to flow between areas only at low inhibition phases [224]. Hence, an α ERD, which is reflected in lower α power, has been interpreted as a proxy for active stimulus processing (stimulus \rightarrow de-correlation of the signal \rightarrow less synchronization \rightarrow lower amplitude after frequency decomposition). Thereby, an activation of neurons caused by the onset of the stimulus is followed by an ERS once stimulus processing is finished, which brings α power back to baseline [230]. Phase and amplitude thereby play a crucial role in modelling the time windows at which an exchange of information can occur. High α amplitudes would shrink the window where information can possibly be transmitted. It has been suggested that inhibitory α oscillations are implemented by bursts of pulsed inhibition [231]. Neuronal population A thereby can only transfer information to neuronal population B in the absence of such a pulse (trough of the oscillation). If the signal from A to B would arrive at the same time as the inhibitory pulse (peak of the oscillation), the signal would be "overruled" by the inhibitory activity and thus A could not exert an effect on B . See Figure 2 for a depiction that coarsely follows the illustration of Figure 1 in Klimesch et al. (2007) [224]. Evidence supporting the GBI hypothesis is twofold: observed ERD (and ERS) in response to certain stimuli and changes in cortical connectivity related to power changes in the α band.

Verification of the GBI hypothesis not only requires the verification of α ERD in task relevant regions, but furthermore requires evidence for α ERS in task irrelevant regions. Händel et al. (2011) [173] demonstrated that α ERD is positively and α ERS is negatively correlated with attention. In this study, participants were cued to attend a particular side of the screen, where with a probability of 80% a target stimulus would occur, which participants were asked to report on detection by a button press. Contra-lateral to the attended side, α power significantly decreased over occipital sensors, while at the same time a significant increase in α power could be observed on ipsi-lateral occipital sensor sites. Those results can be interpreted as direct evidence for the inhibitory nature of α ERS over task irrelevant regions. Additional evidence for the inhibitory nature of α oscillations is provided by research relating perceptual detection performance to the phase of ongoing α oscillations. It could be demonstrated that visual perception performance strongly depends on the phase of α oscillations measured at task related sensors sites [205]. Again, α power was negatively correlated with detection performance. Those findings indicate a strong negative relationship between α band power and the recruitment of cortical areas to successfully perform a set of visual tasks. However, the respective phase dependency is tightly linked to the underlying α power, which indicates that α behaves more in a burst like fashion (short lived creation of a

gradient of excitability) as compared to a sinusoidal rhythm [232]. Furthermore, target facilitation and distractor inhibition linked to α power de-/increases have been questioned recently [176]. Only under conditions, where the distractor directly interferes with the information processing of the target, supposed suppression effects could be reliably observed [233]. A likely explanation is that the GBI hypothesis does not explain *all* α activity. Indeed multiple α sources have been hypothesized to carry out different functional purposes [234]. Some frameworks suggest that (primary) sensory processing related oscillations in the α band and oscillations tied to more attention related process (or processes related to the modulation of information flow [148]) in the same band in higher order areas (e.g. FEF) actually reflect two distinct computational roles [91]. The GBI hypothesis does not explicitly make this differentiation but the proposed mechanism has nevertheless been explicitly tied to attention.

It has been suggested that α ERD in sensory areas of the cortex, occurs as a consequence of anticipatory top-down inhibition of thalamocortical relay cells, forcing them to switch from "tonic" inhibition mode to "burst" inhibition mode, which allows for stimulus timing as described above [224]. This idea receives support from animal studies, suggesting that the pulvinar (a nucleus of the thalamus) influences cortical synchrony in the α band [235]. In this study, monkeys were trained to perform a variation of the Eriksen flanker task. Thereby, a spatial cue indicates the respective location of a target stimulus that is flanked by distractors. It could be shown that pulvinar neurons reliably responded to cues in their respective receptive field. If the monkeys additionally attended the respective location corresponding to the receptive field, the response was significantly greater. Importantly it could be shown that the influence of pulvinar neurons on the V4-TEO phase relationship significantly increased in the time between cue onset and target onset for receptive field locations that were attended. In the light of the GBI hypothesis, those findings can be seen as evidence for an involvement of the thalamus in modulating cortical connectivity based on α oscillations. However, whether the thalamic influence on cortical connectivity (as suggested by Klimesch et al., 2007, [224]), was triggered by a cortical top-down mechanism, has not been tested. Recent work suggests that α oscillations have multiple origins, depending on the task, exact frequency band and recording site [234]. Even intrinsic properties of neo-cortical pyramidal cells, specifically sodium and calcium channels in layer five neurons, have been identified to yield sufficient explanatory power for the origin of α oscillations [236]. It has been further argued that respective results on α as an inhibitor agent could also be interpreted in the light of α not acting inhibitory to suppress distracting stimulation, but instead might act as an enhancer for active processing [176]. If α would act as a simple suppressor in order to hinder active processing of distractors, then there should be no relationship between distractor features and α . However, exactly this relationship could be found [177]. In general, the GBI seems too simplistic in order to explain many recent findings on α and its inhibitory or dis-inhibitory role (see also " α " in "Neuronal oscillations" on page 27).

(see section "Neuronal oscillations"), have been linked to visual feature binding [183, 184] but it has been refuted [186–188]. A transformation of visual input along the visual stream would require feature specific communication within and between each level of the visual hierarchy, in order to extract and process relevant feature and form a coherent percept of the visual input. A simple model for the purpose of illustration would be the Neocognitron (see Artificial neural networks: Not so distant relatives) [134] which directly inspired by findings on simple and complex cells in the visual cortex [30]. In the light of the CTC hypothesis, the Neocognitron model would require a functional path from shallow to deep layer feature spaces that e.g. enables object recognition signalled by some output layer. This path - again, in light of the CTC - would be laid out by synchronized activity of involved neuronal populations. Unwanted (e.g. feature unspecific) input could partly be eliminated since uninvolved neuronal populations would not fire in sync with the target region. The "retinal" input of the Neocognitron would be send to a first processing layer, which would respond to simple features like oriented bars. Specific complex cells (here C-cells), would respond to the combined activation of feature specific simple cells (here S-cells). Hence, C-cells would be specific to a set of *combined* features. Again, in the next layer, S-cells would respond to specifically activated C-cells from the previous layer, and again be combined into even higher order features by another layer of C-cells. In the light of the CTC the path through the different hierarchical levels would be paved by coherent γ oscillations.

Evidence for the CTC hypothesis stems from findings on anatomical connectivity (see also "Functional and effective connectivity" and "Canonical microcircuits: cortical compute modules"), animal studies and studies on humans. The canonical microcircuit model for predictive coding [82], which is often used as reference model for local functional connectivity, postulates that γ band activity could be attributed to feed-forward processes originating mostly in superficial cortical layers, which indeed could be verified by electric stimulation of macaque V1, which elicited a superficial layer γ response targeting V4 [52]. It could furthermore be demonstrated in humans using a selective attention paradigm that induced γ band oscillations indeed reflects coherent signalling between V1 and V4 in feed-forward direction [242]. Additionally it could be shown that γ band coherence lags in monkeys - between sending and receiving region - often range from 3 *ms* (intra-area deep to superficial [243]) to 10 *ms* (FEF to V4 [244]), but could last up to 20 *ms* for connections between inferior frontal junction (IFJ) and the FFA or parahippocampal place area (PPA) [245]. Those findings would clearly speak in favor of a feed-forward directed information flow represented by γ oscillations that reflect synchronized firing patterns for coherent intra and inter-area communication.

While the CTC model provides a straight forward explanation for coherent feed-forward processes through entrainment of post-synaptic gain, feedback processes have not been explicitly incorporated in the model to such an extent. In a later article however Bastos et

al. (2015) [246] explained that the original CTC would threat bi-directional communication as functional network assemblies, formed via (near) zero lag phase synchrony. Thereby β band activity is hypothesized to reflect feedback directed predictions. However, the exact interaction between high frequency feed-forward and low frequency feedback activities is described in less detail. The updated model would instead propose an interaction between feed-forward and feedback directed processes, depending on intra and inter-area γ phase. Each feed-forward flow of information from the thalamus to the visual system would serve as input in L4 from where it is routed towards supra-granular layers (including a small phase delay). Areas along the hierarchy would consecutively modulate the gain of the receiving regions (entrained from the sending region), such that it is phase coherent with the sending region. From supra-granular layers, feedback can now be send to lower order supra-granular layers, since both would be in sync as well. This implementation would yield two anatomically and functionally separable channels for feed-forward and feedback information flow.

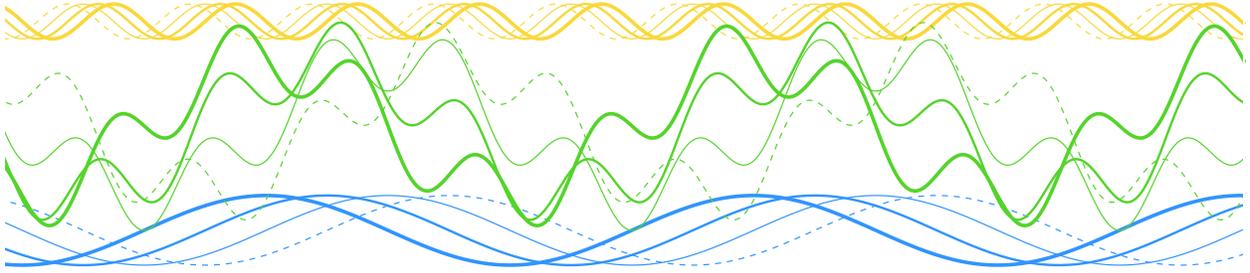


Figure 4: Nested feed-forward high frequency and feedback directed low frequency oscillations. From dashed to more solid and thicker lines represent the respective progression of oscillations over time with the dashed line representing the past and solid line thick lines the present state. Slow (e.g. θ or α oscillations (blue lines) indicate top-down selective processes, whereas high frequency γ oscillations (yellow lines) are associated with the feed-forward flow of information (feed-forward: left to right). The green lines represent the compound frequency of both, where high frequency activity is nested in the low frequency oscillation. See also "Coding via nested oscillations"

Communication via nested oscillations

Models of nested oscillations [247, 248] try to link hypotheses about fundamental principles of cortical information flow derived from findings on slow oscillations like θ or α (e.g. Gating by inhibition), with those about high frequency oscillations like γ (see e.g. Communication through coherence) [240, 249]. Furthermore, anatomical findings about the local cortical structure (see "Canonical microcircuits: cortical compute modules") and functional long-range network architecture (see Structural aspects of the brain) are incorporated to provide the basis for detailed hypothesis testing [91]. Nested cortical oscillations comprise the idea of a compound frequency of multiple base frequencies (see Figure 4), which implement different functionality and are implemented differently within the cortical architecture and frequency spectrum (see also "Coding via nested oscillations"). Thereby "medium slow" oscillations (α , β) are associated with feedback directed connectivity [249, 250]. The model by Bonnefond et al. (2017) [91] explains high oscillations (γ) as feed-forward flow of information reflecting synchronized neuronal spiking [240]. Slow oscillations are considered to facilitate the recruitment of neuronal assemblies (cortical networks), modulated by attention [251] or stimulus anticipation [252]. Neuronal populations involved in ongoing stimulus processing are thought to be combined into a stimulus processing pipeline via (coherent) α ERD. Fast oscillations would use newly established dis-inhibited and synchronized network routes for feed-forward directed information transmission nested in slow oscillations (see also "Coding via nested oscillations" below). Originating from thalamic nuclei, visual information would be projected to granular layers (L4) of the primary visual cortex. The signal is passed internally to supra-granular layers (L2/3), where it is forwarded to L4 of the next target region. Feedback connections - according to this model - originate in supra-granular

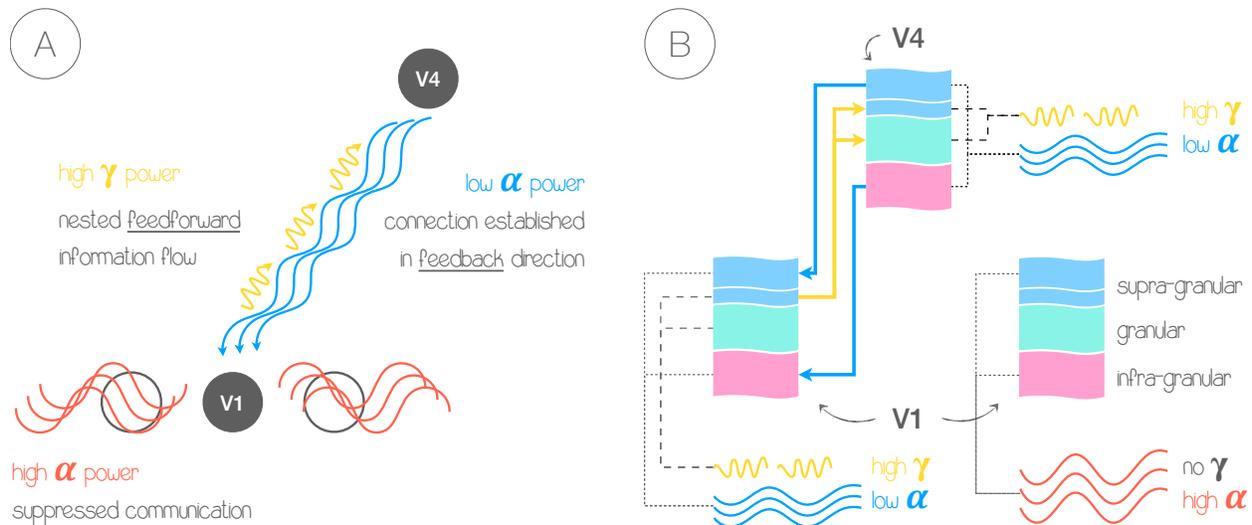


Figure 5: Inter-area communication via nested oscillations. Figure adapted from Bonnefond et al. (2017) [91]. **A) Functional connectivity.** A communication channel between V1 is established in feedback direction (from V4 to V1) by decreased but synchronized α band activity in the respective target areas (blue). High α power in non-target areas suppresses upstream information flow (red). Once a communication channel is established, γ oscillations carry the respective information from V1 to V4 in feed-forward direction (yellow) nested in α (see also "Coding via nested oscillations"). **B) Micro-circuitry of nested oscillations communication.** Communication channels are established in feedback direction by low α power (blue) in supra and / or infra-granular cortical layers (light blue and magenta). The feed-forward information flow is reflected in an increase in γ power (yellow) from supra-granular to granular layers (light blue and cyan) in the upstream population (supra-granular for long range connectivity), nested in coherent but lower power α (blue). High α power (red) - possibly in anti-phase - in V1 supra and / or infra-granular layers interrupts the upstream information flow, suppressing a large part of the respective γ power increase.

and infra-granular layers with respective connection to supra and infra-granular layers of the lower order region (see Figure5 B).

Within the communication via nested oscillations framework, visual information (reflected in the γ band) is transferred along the established communication channel (reflected in the α band) nested in the phase of the low frequency oscillation. Figure 4 illustrates bi-directional information flow based on nested oscillations. Multiple points in time (i.e. snapshots) are represented by line thickness, such that the most thick and solid lines represent near past or present time points, whereas the dashed line would reflect the time point furthest in the past. While the blue line "travels" from right to left, the yellow line travels from left to right. If feed-forward to feedback connectivity is implemented from left to right, the slow oscillation component of the green compound wave travels in feedback direction, whereas the fast oscillation component travels into feed-forward direction. Connectivity in

this model predicts phase coherent γ oscillations in feed-forward direction, originating at superficial to middle layers (L2/3 possibly L1), connected to granular layers of the target region (L4), while - more importantly - slow oscillations are thought to occur prominently in deep and superficial layers, implementing supra-granular to supra-granular and infra-granular to infra-granular feedback connectivity (See also Figure 1). It could be shown that α ERDs ranging from 400 *ms* to 700 *ms* after attention cue onset, are accompanied by a state shift in fronto-parietal attention networks, indicated by a cross-hemisphere lateralization of fronto-parietal functional connectivity [253]. Top-down directed attentional control networks involving the FEF have been widely known for years [254], however the critical involvement of cortical oscillations has been subject to more recent publications. Microstimulation in FEF leads to attention like neuronal spiking in primary visual cortex areas [76], but furthermore this change seems highly related to respective changes in the α band connectivity in top-down (feedback) direction [251]. Findings on feedback connectivity estimates from laminar level fMRI point in a similar direction as well. The perception of an illusory shape has been shown to be related to increased feedback driven deep layer cortical activation in V1 in receptive fields corresponding to the position of the illusory figure [34]. Bottom-up stimulus processing however was reflected by an increase of the overall BOLD signal across all layers, but predominantly in middle and superficial layers. Using this model common findings on e.g. phase coherent γ oscillations (see Communication through coherence) as well as targeted (e.g. attention driven) α (dis-) inhibition (see Gating by inhibition) would reflect two aspects of a joint model of nested processes.

Within the predictive coding framework (see WHAT IS . . . on page 33), feedback directed predictions are reflected in β burst activity [219]. In the light of the present framework, considering functional principles of canonical microcircuits (see "Canonical microcircuits: cortical compute modules"), predictions reflected in the β band could potentially be conveyed downstream by neuronal synchronizations in the α band [219] which results in an upstream directed information flow in the γ band (nested in α) reflecting the prediction error. Since α is hypothesized to target specifically important neuronal populations for the ongoing process, the respective "targeting" would reflect some sort of precision [223] that is set to minimize surprise (prediction error, γ) by conveying predictions (β) to "best suited" neuronal populations, which could increase precision by lowering the false positive rate (falsely active neuronal populations).

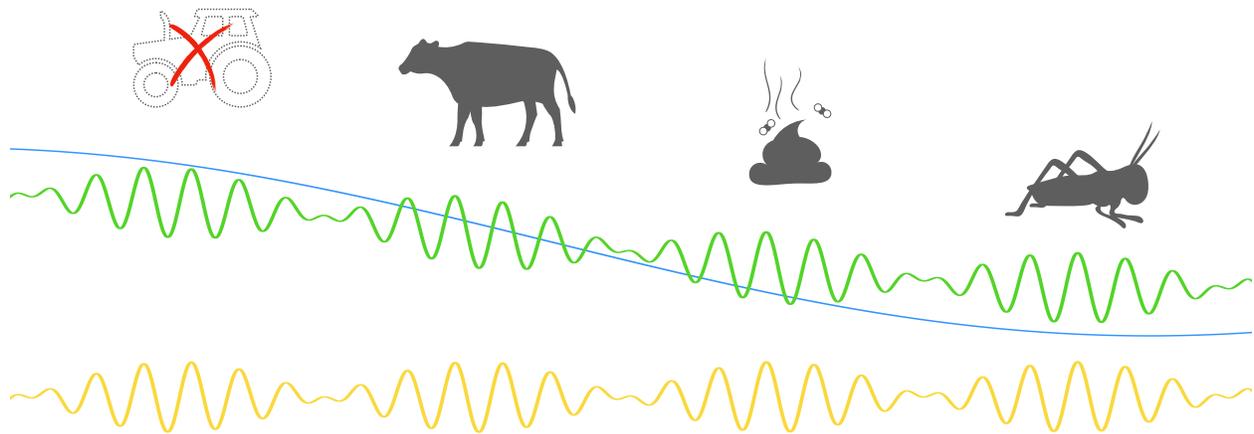


Figure 6: **Multi-item sampling based on nested oscillations** [210, 255]. Perceiving a typical scene from the country side involves the processing of a variety of stimuli of different sensory modalities. Information could be visual (e.g. a cow), auditory (e.g. the sound of a cricket) or olfactory (e.g. the fresh air of the country side), for multi- or mono-sensory information processing. Each item is processed by primary and higher order visual regions, which is expressed as feature [256] or item specific γ oscillations [245]. Each feature specific γ burst corresponding to one item, might arrive from different sensory areas (joint representation by the yellow line). Depending on the underlying level of inhibition - reflected in the α or θ band (blue line) - stimuli (or stimulus features) that evoke a stronger response are processed earlier along the cycle (where inhibition is still quite high). This temporal sorting along the phase gradient allows multiple items (or features) or multi-sensory input to be combined into coherent percepts [257] via γ band oscillations [204]. By modulating frequency or amplitude (not shown here), the "window of opportunity" (i.e. where inhibition is low enough for information to be transferred) can be narrowed (higher frequency or amplitude) or widened (lower frequency or amplitude). Depending on e.g. the number of items that are combined to one "object", low frequency amplitude and / or frequency might vary accordingly (wider window, more features). The green line was created by computing the average between the blue and yellow lines. For illustrative purposes the average, rather than the sum was chosen, such that the blue line (reflecting underlying slow oscillations), could serve a second purpose of modelling the gain of the post-synaptic cell (modulating the threshold at which the post-synaptic cell can be excited). Information is coded along the slow oscillation phase gradient based on the combined nested oscillation activation. The first item-encoding burst of γ which exceeds the underlying θ inhibition (values of green line higher than values of blue line) will be transferred first and successively so forth along the phase gradient for signals exhibiting less excitation. It has been further suggested that the θ phase gradient encodes multi-object, whereas the α phase gradient encodes multi-feature (within one object) information in the γ band [91, 210, 258].

Coding via nested oscillations

Beside the notion of nested cortical oscillations implementing bi-directional information flow, a more functional focused consideration of the duality between low and high frequency oscillations has been discussed as well. Spike time coordination in hippocampal formations

modulated via the γ cycle, nested into the θ rhythm, has been hypothesized to reflect a temporal sorting scheme for multi-item stimulus processing along the θ phase [204]. Low frequency phase dependent sampling has further been proposed a mechanism for attention sampling [207]. Those models are based on the notion of a low frequency phase gradient which allows stronger input signals earlier in the cycle [259]. On the behavioral level, within object target detection performance (computed by varying SOA times) fluctuated around 8 Hz (α band), whereas between object target detection performance fluctuated with 4 Hz (θ band) [210] (see Figure 6). Recent findings from SNN simulations support the idea of sampling based neuronal computations based on low frequency oscillations as well and could demonstrate that oscillatory "background activity" potentially contributes to fast information sampling from multiple sources and speeds up computation time by efficient sampling from the solution space [260].

Local cortical computations and widespread neuronal network dynamics produce oscillatory signals that have been linked to feed-forward and feedback connectivity in the neocortex across multiple cortical layers, spike time coordination in the hippocampus, attention related fronto-parietal network activity and perception related phenomena. Open questions however remain with respect to how functional connectivity on the level of cortical laminae, using non-invasive technologies, such as EEG, MEG, fMRI or network simulations, could be realized. Estimating laminar level cortical activity for multiple frequency bands furthermore provides the possibility to distinguish functionally separate processes that share similar time-frequency properties. Multiple " α s" or " θ s" serving similar computational properties as ingredient for functionally different computations or spatial communication patterns are possible [261]. Connectivity related α ERD - setting up the computational route for stimulus processing - could e.g. be complemented by attention specific α (dis-) inhibition. Indeed thalamo-cortical and cortico-cortical sources for α band oscillations could be identified [262]. Whether those two α loop connections are functionally related to the presumed computational properties mentioned before thereby is less important than the insight that multiple α sources exist. Hence, it is very well possible that e.g. α band activity stems from attention related sources or feature related sources or both. This makes sense, considering how vastly different local and widespread networks are set up, but are operating in the same frequency band. Receptive field specific α (e.g. center surround suppression [263]), feature specific α (e.g. salience dependent phase coding [264]) and attention related α (e.g. to one hemi-field [251]) could hence reflect functionally and anatomically separable processes that make use of an around 10 Hz inhibitory oscillations which in combination implement a complex inhibitory fine grained modulation of the target area (e.g. early visual regions). Functional nesting of e.g. γ oscillations in low frequency phases in feed-forward direction would hence be reserved to connectivity related functional (dis-) inhibition.

1.4 Research question

In a broader sense, this thesis investigates if and to which degree, cortical oscillations play a functionally, computationally and anatomically describable role in the cortical compute architecture. Recent advancements in human electro-physiological neuro-imaging methods, allow for trial based analyses of cortical oscillations that can be narrowed down spatially to the level of cortical layers. It could be demonstrated that using combined EEG-fMRI, differential distributed layer profiles for α and γ band oscillations in primary visual regions can be obtained [238]. Furthermore, connectivity analyses within and between laminar resolved cortical regions, relating BOLD signal connectivity to EEG power spectrum fluctuations, have been proven to be a viable option [265]. Additionally, recent advancements for MEG data acquisition protocols (e.g. using individual head-casts⁹ [267]) allowed the mapping of individual motor related β burst components to a superficial or deep compartment of a two layer cortex model [220]. Furthermore, recent developments enabled laminar level MEG research, spatially resolving two cortical layers (deep and superficial) that could be linked differentially to α and γ oscillations in the visual cortex [268]. Those developments allow to non-invasively test predictions about "Communication via nested oscillations" (see respective section on page 1.3) and anatomical implementations in humans using MEG. Investigating the fundamental compute architecture of the brain on the level of cortical laminae using electrophysiological methods might reveal fundamental network compute principles which can be used to inspire or enrich artificial intelligence (AI) research (see Artificial neural networks: Not so distant relatives). In SNN simulations, the performance of ANNs for certain tasks could significantly be increased by introducing a low frequency "background oscillation", supposedly enhancing the networks sampling-based probabilistic inference capabilities [260]. In addition to that DNN architectures enhanced with feedback directed generators, inspired by the predictive coding framework, exposed a much larger resilience against adversarial attacks [132]. Derived computational models in turn would allow for a refinement of empirical hypothesis testing with respect to theoretical frameworks on the cortical compute architecture.

From animal research and theoretical frameworks about the underlying functionality of neuronal oscillations - in particular the communication via nested oscillations (CNO) model - a set of testable hypothesis can be derived:

⁹A head-cast is used to create a unique, reproducible fit (relative position in space) between the subject's head and the MEG sensors [266]. Individual subject's anatomy - obtained using T1 weighted magnetic resonance imaging (MRI) - has been used to obtain the head-shape. From this, a foam negative is created for the space between the MEG dewar and the subject's scalp. The participants "wear" the head-cast like an individual whole head foam helmet that precisely mounts to the MEG dewar's head opening ().

Hypothesis 1 α power decreases over lower and higher order regions (or neuronal populations) are specifically linked to ongoing stimulus processing.

Hypothesis 2 α oscillations are in synchrony (coherent) between lower and higher order regions (or neuronal populations) that are involved in ongoing stimulus processing.

Hypothesis 3 α power related changes with respect to attention differ from those with respect to ongoing feature processing, which would be expected based on anatomical findings [80] and the idea that those are implemented in separate cortical processes with different laminar profiles [91].

Hypothesis 4 α power decreases are linked to increased activity in neuronal populations that preferred certain stimulus features as compared to the not preferred stimulus features.

Hypothesis 5 α band activity decreases for expected (predictable) stimuli [269].

Hypothesis 6 α power changes are mainly linked to deep (possibly feature processing [34, 52]) and superficial layer (possibly spatial or directed attention [56, 270, 271]) activity.

Hypothesis 7 γ band activity is related to stimulus feature specific processes.

Hypothesis 8 γ band activity increases when predictions about stimulus features are violated.

Hypothesis 9 γ band activity is related mostly to superficial and mid layer neuronal activity for ongoing stimulus processing [52, 91, 238].

Hypothesis 10 γ band activity is nested in α band activity and is coherent between stimulus processing regions [240].

Hypothesis 11 Spatial filters that transform an image into an edge enhanced version can be obtained by training a DNN to reproduce γ band power changes (obtained from a MEG experiment, see below) in the visual cortex from initial stimulus material.

Hypothesis 12 External low frequency oscillatory activity increases robustness against noise in a SNN performing object classification.

In the following sections the respective hypotheses will be elaborated and explained given the respective experimental modality. Thereby, hypotheses mainly linked to anatomical and functional aspects of active sensory processing, are investigated using laminar level simultaneous EEG-fMRI (Hypotheses 1-4, 6, 7, 9, 10). Hypotheses that either require a large number of trials to test (e.g. prediction violation) or mainly focus on temporal aspects, are tested using MEG (Hypotheses 1-10). Lastly, respective results are used as reference data for ANN simulations and theoretical functional principles are implemented in ANN simulations as well (Hypotheses 11, 12). Unfortunately, from the initially planned three-fold approach (EEG-fMRI, MEG, ANN), only the first one was fully executed. This is explained by various circumstances, which will be highlighted in the respective sections (see "Experimental work").

Laminar level EEG-fMRI Computational models involving canonical microcircuits (see "Canonical microcircuits: cortical compute modules") as well as computation and communication models based on nested oscillations (see "Communication via nested oscillations") inevitably rely on empirical testing of specific laminar level oscillatory activity. But only recently respective methodological advancements allowed to conduct laminar and frequency specific human subject experiments using non-invasive techniques on healthy adults [220, 238]. In a first step it needs to be demonstrated that the novel methodological possibilities indeed allow for in depth hypothesis testing about the role of (nested) oscillations in the cortex. Scheeringa et al. (2016) [238] demonstrated in a simultaneous EEG-fMRI experiment that cortical oscillations can be related to the laminar level BOLD signal in primary visual regions. In a detection task, where participants were asked to respond to a speed change in the concentric moving circle pattern, EEG activity in the α band was found to be correlated with the BOLD signal in deep and superficial, whereas the γ band trial-by-trial power changes correlated with the BOLD signal mostly in middle and superficial layers. Since this study provides only a proof of concept for the applicability to functional hypothesis testing with respect to brain oscillations on the level of cortical laminae, applying the methodology to task related hypothesis testing remains to be done. Since occipital γ band oscillations have been related to visual feature related processing [182, 184, 189, 240] predominantly in supra-granular layers [52, 238] where feed-forward pathways to L4 of higher order regions have been located [80], a feature selective γ band response for each level of the visual cortical processing hierarchy would very much be expected (Hypothesis 7). Since cortical neurons in early visual regions respond predominantly to simple features, such as bar orientations [30], the visual response in the γ band for distinct low level visual features (e.g. orthogonal oriented grating stimuli) would be expected to be correlated differential with the BOLD signal, obtained from different neuronal populations in V1 with respective orientation preferences. If indeed γ band neuronal activity reflects feature selective ongoing feed-forward activity, then γ band power should be correlated with the BOLD signal only

in voxel that capture the separation between both stimulus orientations best. Hence, high values for the difference between both orientations reflect in each voxel the preference to one orientation and at least the ignorance (if not suppression) of the respective other. Feature specific activity has been mainly associated with superficial and potentially middle layers [238], but also deep layer activity would be expected to some extent [52, 80] (Hypothesis 9).

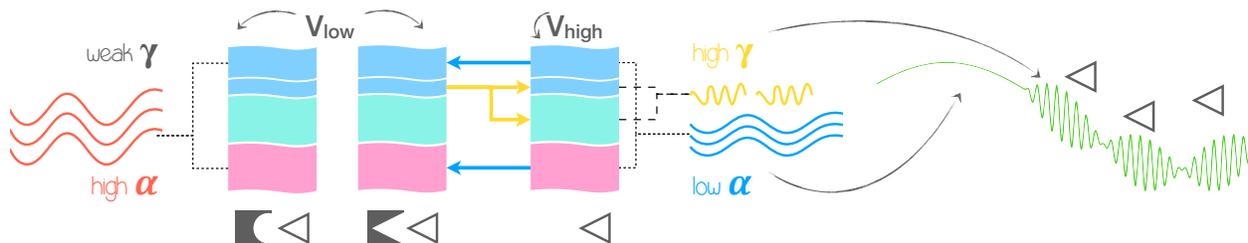


Figure 7: **Feature specific neuronal oscillations.** Using a combined EEG-fMRI experiment, feature specific contributions of high and low frequencies can be related to specific neuronal populations via the BOLD signal. Stimulus features (represented by the triangle), are expected to elicit higher BOLD responses in voxel preferring the respective stimulus feature (square with negative triangle) compared to those who do not (square with negative semi circle). α power in stimulus unspecific or non-preferring regions is expected to be correlated negatively with the BOLD signal in that regions. No γ increase is expected, due to high α inhibition. The BOLD signal in neuronal populations preferring a respective stimulus feature are as well expected to expose a negative correlation with α power. In that case a positive γ correlation would be expected as well. α power effects are expected to be mainly found in deep and superficial layers, whereas γ effects are expected to be associated mostly with superficial and middle layer activity (but possibly also deep [80, 272]). Furthermore, the nested oscillations based information transfer between lower and higher order regions (lower order visual region: V_{low} and higher order visual region: V_{high}) is expected to be reflected in BOLD signal feedback connectivity that is negatively correlated with α power in deep and superficial layers between regions and feed-forward connectivity between regions from superficial to middle layers that is correlated with γ band power. See Figure 4 for a depiction of a bi-directional compound oscillation where one oscillation is nested into a second oscillation.

As Scheeringa et al. (2022) [265] in their proof of concept further demonstrated, oscillatory activity can be related to intra-area (inter-layer) or inter-area BOLD connectivity measures. This gives rise to the opportunity to relate feature specific cortical activity to functional connectivity in the visual cortex. Here, γ band activity would be expected to be significantly related for feature specific voxel connectivity from superficial in the source to middle layers of the target region (Hypothesis 10). From connectivity related findings on laminar fMRI [34], as well as electro-physiological findings from monkeys [52], it can be derived that EEG α band activity should be related to the layer specific BOLD signal in mostly deep, but also superficial layers (Hypothesis 6). This proofs particularly difficult, since the SNR in deep cortical layers is expected to be lower than in superficial layers due to superficial draining veins that have an effect on the BOLD signal [273]. However, if respective α effects expose higher correlation values for deep layers or can be compared to other layer profiles

(like Scheeringa et al., 2016, [238] did) and prove to be different, then the influence of the SNR bias towards superficial layers has only little effect on respective conclusions that can be drawn. Expected deep and superficial α decreases are predicted to be related to top-down feedback connectivity, which again could be tested relating laminar level BOLD connectivity to the EEG power changes in each trial (Hypothesis 2). Whether changes in the α band are indeed related to feature specific populations of neurons (e.g. dis-inhibiting voxel that preferably respond to stimuli in one orientation and inhibiting voxel preferring the other orientation) [91] can easily be tested by comparing the α band correlation of the BOLD signal between voxel exposed to their preferred or non-preferred orientation (Hypothesis 4). Specifically, an α band decrease linked to BOLD activity of neuronal populations preferring a respective orientation would be mostly reflected in deep and superficial layers. Given that α band activity has been associated with spatial attention, but also local canonical microcircuit computations, superficial and deep layer α band negative correlations with the BOLD signal might reflect distinct feedback processes (Hypothesis 3): establishing feature specific communication channels and general attention inhibition / enhancement respectively. For neuronal populations that do not prefer a respective orientation, a spatial attention related α band power decrease would be expected (Hypothesis 1) as well (mostly in the deep layers), whereas a relative α increase (compared to the preferred orientation neuronal populations) would be expected to reflect less stimulus related activation in that population. A negative correlation between α band power and the BOLD signal for neuronal populations that do not prefer a respective stimulus orientation would hence be expected but furthermore that respective (de-) activation patterns between preferred and non-preferred orientation significantly differ. Receptive field specific α power changes can be assessed by comparing the general BOLD activation of voxel responding to any stimulus with an increase of activation, with those responding with a decrease. Since α power was found to be center-surround [263] and attention hemi-field [251] specific, a decrease in α power for activated (dis-inhibited; see "Communication via nested oscillations") and an increase in α power for deactivated (inhibited; see "Gating by inhibition") voxel would be expected. Again, deep and superficial layers are expected to be source (higher order region) and target (lower order region) of such oscillatory communication, where superficial α decreases are thought to reflect stimulus processing related feedback activity and deep layer α more general attentional processes [91]. Lastly, the contrast between preferred and not-preferred feature conditions would shed light onto the differential role of cortical oscillations in feed-forward and feedback pathways across different cortical layers.

The described study has been conducted to a large degree. For a detailed description see the section "Feature specific neuronal oscillations in cortical layers (in prep)" and Figure 7 for a visualization of the respective research question. A respective publication is in preparation as of today (January 15, 2023).

Laminar level MEG A major issue in laminar level EEG-fMRI is the low trial count per experiment (due to the time it takes to even collect a single fMRI volume) and the relatively low SNR, considering that neuronal fluctuations in the range of ms - obtained from a blurred scalp EEG signal - are used as predictors for BOLD signal activity changes in the order of s - obtained at very precise defined anatomical regions. While the latter issue can partly be overcome by precise modelling of the hemodynamic response function [238], the limited trial count per subject limits the complexity of experiments that can be conducted using this method. As Bonaiuto and colleagues (2021) [220] successfully demonstrated, the mapping of specific aspects of neuronal oscillations (β burst components in the motor cortex) to two cortical layers is achievable using MEG irrespective of the inherent spatial constraints imposed by this method. Furthermore, laminar MEG has been shown in a proof of concept to be suitable for laminar level (time-) frequency level inference in the visual cortex in general [274]. Since laminar level MEG does not rely on the BOLD signal to generate required spatial precision, a more direct link between cortical activity with respect to neuronal oscillations becomes accessible. Due to the potentially massively increased number of trials by conducting a MEG experiment over fMRI, more complex phenomena, such as spatial attention, top down predictions or brain network connectivity become viable for non-invasive studies on healthy human participants allowing to investigate laminar level brain oscillations. Nonetheless, fMRI data can be easily related to anatomical structures, while laminar level MEG relies on a multitude of prior assumptions to ensure precise spatial mapping [267]. Since, feedback directed communication has been associated primarily with deep layer activity [34, 52] and feed-forward activity with superficial layers [52, 238] a two layer model might be sufficient to differentiate low and high frequency oscillations across the visual stream in primary visual regions and higher order areas. Additionally, recent methodological developments in our department with respect to individual head-casts - mainly led by James Bonaiuto [266] and Denis Schwartz - in combination with a supine (lying down) positioning of the subject inside the MEG have led to promising preliminary results that suggest the feasibility of even resolving three cortical layers.

In order to assess the relationship between stimulus features, cortical processing and nested neuronal oscillations within the realm of respective frameworks at the resolution of cortical layers (Hypothesis 1-10; detailed out in section "MEG experiments" on page 117), the development of suitable experimental setups and tasks is crucial. In general the targeted effects are expected to suffer from low SNR on the attempted spatial resolution due to the physical constraints imposed by MEG as a neuro-imaging tool. Hence, ideally respective tasks must be designed, such that a large number of trials can be collected as well. To optimize for different aspects of feature processing, two MEG studies were planned for this PhD project. In Experiment 1 (see Experiment 1 on page 123) more local aspects related to canonical microcircuit models are tested, whereas in Experiment 2 (see Experiment 2 on page 130) feature selective attentional processes as well as long range hierarchical stimulus

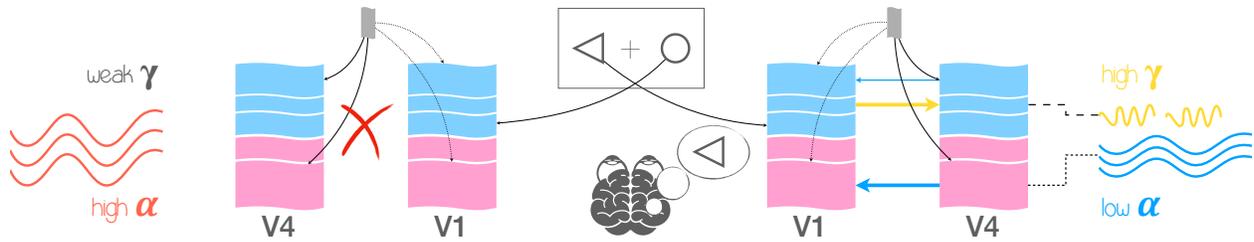


Figure 8: **Top-down modulated neuronal oscillations.** Using laminar level MEG high and low frequency modulations in primary visual cortex could potentially be mapped to deep and superficial layers (based on methodological developments for β bursts [220] and a proof of concept for the visual domain [268]). This extends the range of possible non-invasive methods for measuring cortical layer specific oscillations from combined EEG-fMRI [238] to laminar level MEG for research on the visual system. The figure depicts the expected cortical activity for a subject exposed to a set of two stimuli (one in each visual hemi-field). One stimulus (the triangle) appeared at the cued location (as expected) and exposed the target shape (triangle shape). Cortical areas corresponding to receptive field locations of the unattended hemi-field are expected to expose an α increase. A violation of that prediction (e.g. the target shape appeared on the unattended side), would be reflected by either α phase dependent missing of the target [207, 210] or a detection of the target, which would imply a switch in hemispheric α activity. Within the processed hemi-field (contra-lateral hemisphere), lower α power in deep and higher γ power in superficial layers would be expected [52, 238], reflecting feedback and feed-forward connectivity respectively. Potential higher order top-down influences (e.g. from FFA or visual word form area (VWFA) in a face / word Stroop task) on V4 or V1 potentially expose a similarly complex network behavior as found for the influence of dorsal visual cortex on areas in the ventral stream [275]. Furthermore, the dorsal attention network [276] and pulvinar [276, 277] potentially modulate respective stimulus specific activity [91]. It is hypothesized that stimulus related α band decreases reflect feedback directed communication, where coherently oscillating neurons exchange information reflected in the γ band nested in α .

processing is targeted. In the first experiment feature specific processing is reduced to a minimum (gratings) but stimulus predictability has been modulated, such that predictions derived from canonical microcircuit models can be targeted (see "Canonical microcircuits: cortical compute modules" and WHAT IS ... on page 33). In the second experiment, long range feature selective attention modulated responses in the α range are addressed. To avoid stimulus confounds, compound stimuli that expose at least two conceptually different feature aspects without changing the retinal image must hence be used. Those conditions are met by Stroop stimuli¹⁰ [280]. Higher order features (e.g. words and colors or in this experiment words and faces) compete over interpretive sovereignty and are expected to elicit similar primary visual area response patterns in low and high frequency oscillations, however the α driven network communication should vary between FFA [281] or VWFA [282] depending on

¹⁰The Stroop effect [276, 278], is a psychological effect that describes a delay in reaction time, for feature in-congruent stimuli. Classically, participants would be asked to e.g. name the font color of in-congruently colored words: RED GREEN BLUE. However, any conflicting stimulus could serve as a Stroop stimulus. For example the relative length of the word LONG could be asked or the task entirely modified to an emotional version, where emotion related words and pictures (mis-) match [279].

the feature that was attended. Thereby, γ band oscillations, nested into respective feedback directed α phases, would be related to feed-forward connectivity between regions involved in the respective processing. Similar to dorsal and ventral visual pathways for "perception and action" ¹¹ [283], it is expected for complex feature Stroop tasks that the visual processing stream at some point separates to target respective high level features (e.g. faces or words in FFA or VWFA), which is hypothesized to leave a trail of cortical activity, similar to what has been observed in monkeys [52, 272]. For feature competing stimuli (e.g. complex Stroop), a respective increase in α power in the non-target region is hypothesized to facilitate visual separation and hence should be linked (in phase and / or amplitude) to behavioral performance [91, 210, 251]. Popov et al. (2018) [275], found increased top-down connectivity from dorsal visual cortex (dVis) to V1, V2, V3, ventral visual cortex (vVis), MT and medial temporal gyrus (MTG) in the α band, which was hypothesized to reflect the engagement of multi-region brain networks during a working memory task. If similar network communication patterns could be narrowed down to functionally well defined regions (e.g. FFA or VWFA) at the level of cortical laminae, theoretical computational models based on the cortical architecture can be put to the test.

On a local level, proposed oscillatory properties described by models of canonical microcircuits (see Canonical microcircuits: cortical compute modules) [82] could be studied to a great degree of anatomical precision as the mapping of β burst components to a two layer cortex model [220] using laminar MEG indicates. In a first step, findings by combined EEG-fMRI (see section "Feature specific neuronal oscillations in cortical layers (in prep)" or Scheeringa et al., 2016, [238]) would need to be replicated. A two layer cortex model, would give rise to the predictions that γ band ERSs can be observed in the superficial layer as a response to grating stimuli, whereas α power would be expected to decrease predominately in the deep but potentially in the superficial layer as well. Relating those findings to a cognitive task (e.g. attention), would furthermore provide potential insight into the cortical compute architecture in early visual regions. By introducing a lateral distractor stimulus in addition to the target stimulus, hemi-field specific α ERDs (contra-lateral to the target) and ERSs (ipsi-lateral to the target) [251] effects would be expected. Top down attention control from V4 to V1 has been observed in deep and superficial layers in macaque monkeys and has been related to γ increase and α decrease in those regions [272]. Spatial (covert) and stimulus feature related attention processes are hence expected to reflect separate neuronal processes that can be decoded based on the frequency response patterns even within a similar frequency spectrum. Core predictions of the nested oscillations model by Bonnefond et al. (2017) [91], such as feed-forward γ between e.g. V1 and V4 in superficial layers, nested into feedback

¹¹It has been hypothesized that the visual stream separates into a dorsal stream towards parietal regions which translates visual input for sensori-motor guidance (the *where* stream) and a ventral stream towards temporal areas where object recognition is performed (the *what* stream) [283].

coherent α oscillations between V4 and V1 from deep to deep and superficial to superficial layers, as well as long range connections from frontal areas in the α band [251] can be tested. Thereby, the respective phase relationship between source and target region in the α band as well as nested feed-forward γ , stimulus and region specific, signalling can be further investigated. In the past it has been demonstrated that deep layer top down influence modifies response patterns of neuronal populations in primary visual cortices. Kok et al. (2016) [34] showed that the illusory shape of an optical illusion would cause feedback related laminar fMRI activity in deep layers, which vanishes after slightly rearranging the figure, such that principle perception is not altered, but the illusory shape disappeared. Within the predictive coding framework [131], the brain complements sensory information with internally generated predictions about the current state of the world that are compared against each other. Thereby, using an unknown function, predictions and prediction errors (mismatch between prediction and evidence) can be mapped to each other. Bastos et al. (2012) [82] proposed that the frequency spectrum of cortical predictions, encoded by deep layer neurons, is biased towards low frequency oscillations, whereas the spectrum for prediction errors, encoded by superficial layer neurons, is biased towards high frequency oscillations, due the mapping process between prediction and prediction errors. Hence, depending on the predictability of the stimulus, the size of the prediction error and the relative contribution of low and high frequency oscillations in deep and superficial layers are expected to modulate β and γ band power differentially. Again, respective neuronal communication between neuronal populations would be achieved by coherent α activity. Superficial γ power would be expected to increase depending on stimulus predictability and attention modulation, depending on the hemi-field. Conversely, modulating predictability would be reflected in modulations of low frequency (possibly β [82]) oscillations in deep layers and feedback direction. With respect to the nested oscillations framework (see "Communication via nested oscillations") it is predicted that during anticipation of the stimulus, feedback predictions (expressed as β bursts) are nested in coherent α activity, conveying them downstream. Again, incoming information (after stimulus onset) would be compared to the prediction and respective errors (eliciting activity in the γ band) are carried feed-forward, nested into coherent feedback α . This in turn would result in an update of the predictions which establishes a feedback loop that possibly spans across the entire visual hierarchy.

Artificial neural networks The trilogy between theoretical models (see e.g. "Communication via nested oscillations") empirical evidence (see e.g. above or "Feature specific neuronal oscillations in cortical layers (in prep)") and replication via simulation often spawns even complete lines of new research. Inspired by early findings on computational properties of the visual cortex [30] that developed into artificial neural network models [73, 134], the foundation for modern day research on artificial neural networks was laid. While state of the art DNNs outperform humans in a variety of tasks, their biological plausibility is limited

and their power consumption and training data requirements immense. Conversely, spiking neural network models are much more biologically plausible, but could be computationally very expensive [103, 104] and are much harder to train for certain tasks, but potentially are more sparse by design [284]. By definition, the brain is a spiking neural network. Nevertheless, it could be shown that classical DNN layer operations, such as convolution, can be performed using SNNs as well [285, 286]. Converting DNNs to SNNs however is not a straight forward task. Floating point activation values of the DNNs cannot be mapped 1 : 1 to a spiking neural network. Rate based models (converting the spikes of output neurons to a rate per time) solve this issue by evaluating an extensive time period to *some* extent. In the attempt to optimize this problem, it was found that rhythmic inhibition reduces computational cost dramatically (compared to DNNs), while maintaining output accuracy [287]. Since weight matrices of DNNs can be - to some extent - copied to SNNs (implemented as varying synaptic strength) [288], the opposite conversion from neural network time resolved data to DNN layer weights has been hypothesized to carry meaningful information about the underlying SNNs functional architecture. Indeed, Seeliger et al. (2021) [124] were able to demonstrate a potential solution for the conversion from fMRI data to a set of DNN layer weights. Almost 24 *h* of fMRI recording data has been obtained from a single subject while watching naturalistic stimuli (a TV series). A generative DNN model - set up to mimic the early cortical visual processing architecture (V1-3, MT, FFA) - was trained on a reduced version of the original input that was presented to the subject, to generate the fMRI signal obtained for each region of interest (ROI). Thereby, the loss function was computed directly from the univariate predictions and the BOLD signal in a specific region. After training the model, feature maps of trained layers in the artificial V1 area have been extracted. Feature maps learned by the convolutional layers contained to a large degree oriented Gabor patches of varying spatial frequency, many of which contained a temporal component as well. This approach could be adapted to MEG as well. So far it has been demonstrated that the MEG signal for viewing of naturalistic objects and DNN response patterns correlate more the more time after stimulus onset passes and in layer correspondence (early after stimulus onset only primary visual areas and shallow layers of the DNN correlate, whereas late, deep layer activity and higher order brain regions correlate stronger) [126]. A similar relationship has been found for the relationship between MEG and DNNs in the linguistic domain [289]. Applying novel methodological developments to achieve laminar resolution for MEG (see above or Bonaiuto et al., 2021, [220]), enables the training of DNNs on superficial and deep layers MEG signals respectively. Combined with the aforementioned developments in brain-to-DNN feature mapping, laminar MEG could be used to train a model to generate temporal feature maps (similar to spatial maps like Gabor patches, however rather than mapping space to space, mapping space to time).

Instead of deriving potential candidates for cortical functions from the fit between DNN models and the data, which as been shown to have some caveats [128, 130], SNN models

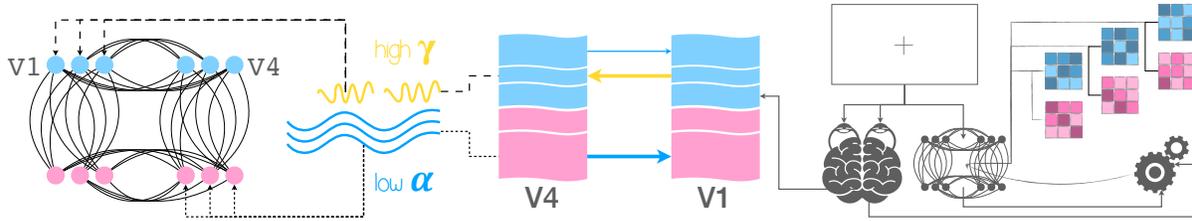


Figure 9: **Brain inspired artificial neuronal networks.** Artificial neural networks can contribute to the investigation of the role of neuronal oscillations in multiple ways. On one hand, biologically derived SNNs (left side of the figure), can be built to investigate the conceptual role of neuronal oscillations. It has been shown that SNN models exposed to a background oscillation benefit from this oscillation in sampling based computations [260]. In that spirit, feed-forward and feedback connected SNNs could be built around the predicted cortical and rhythmic architecture derived from canonical microcircuits and communication via nested oscillations models (see "Canonical microcircuits: cortical compute modules" and "Communication via nested oscillations") and compared in terms of performance and sparsity to other neural network architectures and in terms of network activity to human brain data. Furthermore, it has been demonstrated that DNN weight matrices learned from functional brain data can give rise to meaningful interpretations about cortical "weight matrices" [124]. Even temporal dynamics have been successfully mapped between hierarchically organized DNN models and the visual cortex hierarchy in human participants [126]. Since laminar level MEG (see above) pushed the boundaries of spatial resolution for non-invasive electro-physiological methods, spatio-temporal patterns could be learned from the MEG data (Hypothesis 11), by using a generative DNN that learns to predict MEG data from input stimuli (right part of the figure). Idealized and simplified in this figure, the temporal development of spatial filters in one channel (pink) proceeds with a different change rate over time as compared to the other (blue). Mapping out spatio-temporal weight matrices of the DNN might provide insight into spatio-temporal dynamics of the brain.

that implement cortical functionality derived from theoretical frameworks and experimental evidence, might provide a powerful tool for cortical function mapping in general (Hypothesis 12). Recent experiments on oscillation based SNNs show improved generative properties and sampling [260]. The authors of the study argued that oscillatory background activity can be seen as equivalent to simulated scheduled cooling, where temporary temperature rises reduce the risk of getting stuck at a local attractor. Oscillatory background activity has been argued to implement sampling based computations similar to brain networks. A similar model could be extended to a compound wave background oscillation (see Figure 4) or low frequency feedback and high frequency feed-forward activation in different layers. Parameters could be tuned by relating actual laminar MEG data to SNN activity. From synaptic weight matrices and the respective temporal activity, spatio-temporal activation patterns can be computed. Similarly, spatio-temporal connectivity measures from the SNN can be related to the MEG signal. Furthermore, high resolution time-frequency resolved MEG data could be used to train a generative model, similar to what has been done in the domain of fMRI [124].

Introductory conclusion Laminar level EEG-fMRI, as well as laminar MEG and artificial neural network simulations provide powerful tools to investigate local computations and widespread network dynamics with respect to the role of cortical oscillations. In a rapidly changing environment, efficient processing of sensory information and controlling targeted behavior has been unlocked by a large variety of species who have developed neural networks [6]. As non-stationary organisms, animals over plants require much faster environmental adaption protocols than sole evolutionary processes or simple chemical reaction chains [7]. Since neural networks are expensive compared to other body parts of similar mass [16] the resulting emergent behavior must be capable of at least compensating the additional energy uptake. Indeed, the human brain consumes only around 20 *Watt/h* of energy [17], which given its performance and compared to silicon based hardware is remarkably little [18]. Learning about the computational properties of the brain hence unlocks potential access to a very effective compute architecture, which could help the development of better artificial intelligence systems [290]. Furthermore, understanding the exact computational properties in the human brain, can help to improve brain implants (e.g. deep brain stimulation [19], retina implants [20], cochlea implants [21], brain controlled prosthetic limbs [22]). Theories about brain network communications and local computations often include neuronal oscillations (see "Neuronal oscillations") as computationally relevant aspects of neuronal activity to coordinate sensory information flow [54, 91, 148]. Based on data recorded from monkeys or intra-cranial recordings from humans, those models predict α band related feedback activity mainly located to deep cortical (but potentially also superficial) layers. Inter-area connectivity would be achieved by a feed-back dis-inhibition of a computational pathway reflected in an α power decrease for respective neuronal populations and feed-forward γ band activity which - nested into the lower frequency - reflects feature specific processing. New methods developed to map brain oscillations to cortical laminae, such as laminar level combined EEG-fMRI [238] or MEG [220] became recently available, such that layer specific computational models can be tested non-invasive in healthy humans. Laminar level EEG-fMRI can be used to separate feature and receptive field specific activity for high and low frequency oscillations (see "Feature specific neuronal oscillations in cortical layers (in prep)") across cortical layers and potentially enables laminar connectivity within and between cortical areas with respect to α and γ band power [265]. A two layer cortex model could be used in combination with precise source localization methods up to the level of cortical laminae using MEG [220]. This gives rise to the opportunity to test a greater variety of cognitive tasks compared to fMRI studies and the respective limited trial count. Covert spatial as well as feature specific attention is expected to be reflected in cortical connectivity increase in the α band in top-down direction, targeting mainly deep layers and an increase in the γ band in superficial layers reflecting feed-forward connectivity (synchronized by α , see "Communication via nested oscillations"). Lastly, the development of artificial neural network models, in the attempt to replicate "cortical behavior", provides a veritable method to put computational theories to the test. This can be achieved, by mapping the

input space to cortical activity using a DNN in order to extract possible set of weights that gave rise to that activation [124], or by implementing biologically plausible SNN models that are perturbed periodically (e.g. by a nested oscillation) to reveal computational properties of rhythmic fluctuations [260].

2 Experimental work

Methodological developments made up a core part of the present work. This comprised both, software and conceptual developments. To answer each of the respective research questions (see "Research question") custom analysis pipelines and / or non-standard experimental setups and / or stimuli needed to be created for EEG-fMRI and MEG experiments as well as ANN simulations. In the spirit of open source, those developments are made or will be made publicly available and open access versions of papers are planned or already available. In this section methodological and software developments that had been made in order to answer the respective research question are presented separately for each of the thematic blocks (EEG-fMRI, MEG, ANN). Since the EEG-fMRI publication almost reached submission status as of today (January 15, 2023) most of the respective developments are presented in the quasi-publication-styled section "Feature specific neuronal oscillations in cortical layers (in prep)" on page 58. Furthermore, additional developments that are not included in the future publication but nevertheless are considered important developmental steps for this experiment are shared as well. For both MEG experiments and the ANN simulations the current state of development will be discussed. Again, respective sections have been arranged roughly according to a more or less publication default style. A short introduction ensures conceptual embedding of the methodological developments that make up the core part. Since no results with respect to those MEG experiments have been obtained personally by myself yet, some preliminary results kindly provided by colleagues are included as well together with a respective discussion.

2.1 COVID 19 remark

The COVID 19 crisis turned out to be a major setback for the present project. Especially the planned MEG experiments (see "MEG experiments" on page 117) suffered from lockdown and hygiene regulations. Our MEG department was closed for more than six months which caused a substantial delay to the recording schedule. This fact, combined with unexpected delays in the ethics approval procedure delayed the recordings by more than a year. This means that within the realm of the present thesis, only the conceptualization and implementation of tasks as well as the data collection could be achieved. Results from the MEG experiments are relatively limited. Furthermore, home office regulations as a result of the global pandemic affected me personally very much, due to a very difficult housing situation.

2.2 EEG-fMRI experiment

Feature specific neuronal oscillations in cortical layers (in prep)

Clausner, T., van Mourik, T., Marques, J., Haak, K., & Scheeringa, R.*, Bonnefond, M.*
Feature specific neuronal oscillations in the α and γ are differentially linked to cortical layers.
(in prep)

* both authors contributed equally to the project

The respective processing pipeline will be set to publicly available once the paper has been submitted at: <https://github.com/TommyClausner/laminarfMRIv2>. As of today (January 15, 2023) this has not been the case.

Abstract

Cortical rhythmic activity patterns have been related to a variety of cognitive processes. Their particular computational role has been a long debated topic that resulted in a variety of theoretical frameworks used to explain past findings. Thereby, α band activity has been related to inhibitory processes of attention, affecting sensory processes as well as other cognitive functions. In turn γ band oscillations have been related to a more direct measure of ongoing neuronal activity and is thought to be related to stimulus feature processing. Current theoretical frameworks rely on the separation of the cortical architecture into layers to explain findings mostly derived from animal models. Only recently methodological advancements allow to test layer specific cortical oscillations allowing in depth hypothesis testing of respective frameworks non-invasively in healthy human participants. The here presented simultaneous EEG-fMRI experiment investigates oscillatory activity in the visual cortex for the α and γ band as a response to two orthogonal stimulus features. The most striking result we found is a layer and frequency dissociation within the α band for attention-related and feature related processes. We found that general activation (attention effect) related α oscillations ($8 - 10 Hz$) were negatively correlated mainly with BOLD signal in superficial layers. As predicted by some frameworks, but not by the majority of the literature, we found that α oscillations, in a higher frequency band ($11 - 13 Hz$), to be specifically related to stimulus feature specific BOLD signal in deep and superficial layers. More interestingly, both effects (low and high frequency α) were also observed for negative BOLD signal in line with the inhibitory role of α oscillations. Finally, as predicted, γ oscillations were positively correlated with feature-specific signal in superficial and deep layers.

Introduction

Whether brain rhythms have specific computational roles remains a hot debate in the literature. One of the most popular frameworks suggests that γ rhythms (> 40 Hz) would be involved in setting up specific communications between populations of neurons [240]. This framework further implies a role of feedback related α / β in modulating this communication. Other frameworks emphasised the role of α oscillations in controlling the flow of information within networks [91, 291]. Communication would be set-up via the synchronisation of feedback controlled α oscillations between relevant pools of neurons. Stimulus induced γ oscillations would be controlled locally by these α oscillations, i.e. nested within their excitable phase, and transferred to the next hierarchical level for further processing [91]. Moreover, as α oscillations might be related to functional inhibition, high α amplitude in irrelevant pools of neurons might as well be crucial to ensure the specificity of communication [91, 224, 231, 251]. Beyond this idea, inhibition of non-specific pools of neurons for enhancing the specificity of local processing of stimuli has been suggested as well [173, 292, 293]. These different mechanisms suggest the co-existence of multiple α sources. Communication through α synchronisation implies that α oscillations are at least receptive field specific [176] if not feature specific during stimulus processing. Such a finding would go against the current idea regarding the spatial selectivity of α . Instead, α oscillations related to inhibition should express higher power in irrelevant pools of neurons than in pools of neurons processing a relevant stimulus.

Across cortical layers, different distinct neuronal projections have been found [294]. Those have been linked to a multitude of frequency bands in animal studies [52, 295–297]. Furthermore, simulation studies have confirmed that the origin of those cortical oscillations are potentially linked to the cortical architecture itself [298]. For human studies however, the relationship between the cortical architecture and different frequency bands remains less investigated, mostly due to technical and ethical challenges. Previous work has focused mainly on methodological obstacles [80, 220, 238, 265] or relied on laminar level fMRI only [34, 299–302]. Hence, it remains widely unclear how low and high frequency oscillations and related models [91, 224, 240] are reflected in the laminar architecture of the neo-cortex in healthy humans. The different frameworks suggest that γ oscillations should be feature specific (see also Scheeringa et al., 2011, [303]), i.e. high in neurons specifically associated with the stimulus processed (or with prediction error; see Schneider et al., 2021, [304] and Sedley et al., 2016, [219]). Derived from anatomical findings [80] and studies on animals, some predictions with respect to the activation profile for α and γ can be formulated. If α oscillations are related to feedback processes from higher order regions [52, 219, 242] or the pulvinar ([235, 305]; but see also Saalman et al., 2011, [306]), they are expected to be found across all cortical layers [271, 307], but possibly a dissociation with between general or spatial attention and stimulus feature related α is to be expected as well [91]. For γ oscillations,

a clear relationship between increased post-stimulus γ power and at least feature processing related BOLD signal changes (if not even stimulus feature specific processing [189]) in superficial (and middle) layers [52, 91, 238] are expected and have been found. Findings from animal models [52, 272] and anatomical constraints [80] imply an additional deep layer related signal transmission component that is possibly located in the γ band [246] and has been linked to more spontaneous synchronizations [308] as compared to feature carrying coherent γ oscillations in superficial layers [53].

In humans, electroencephalography (EEG) or magnetoencephalography (MEG) record a mixture of these different signals with a relatively low spatial resolution (but see Michalareas et al., 2016, [242] or Bonaiuto et al., 2018, [268]). However, combining EEG recordings with high-resolution (laminar-level) functional Magnetic Resonance Imaging (fMRI) allows to contravene these issues [238, 274] and test for the aforementioned hypotheses. In the present study, we used such an approach for the first time, focusing at the visual cortex while participants were presented with left or right oriented gratings. We were able to extract feature specific BOLD signals in voxel encompassing three compartments associated with deep, middle and superficial cortical layers in V1, V2 and V3 regions. Interestingly, we further found that negative BOLD signals also partially exhibited feature specificity. In addition - as expected from the literature - we could demonstrate that γ oscillations were correlated with feature specific BOLD signal over deep and superficial layers [52, 52, 91, 238, 272]. More strikingly, α band oscillations could be related to a more general - possibly attention driven - process *and* a differential feature response. Since, a positive BOLD signal will reflect difficult to separate feature and attention related processes, respective feature selective signals have been used as control regressor for the general *deactivation*. Negative correlations between α oscillations and the BOLD signal have been found from 8 Hz to 14 Hz for positive and BOLD signals as well as the negative BOLD signal with feature specific BOLD (\pm) as control regressor, predominantly in superficial layers. We could show that this pattern is different from more feature related α band oscillations. Thereby, an inverted frequency profile for low and high α has been observed to specifically be negatively correlated to voxel preferring (11 – 13 Hz) or not preferring (8 – 10 Hz) a respective stimulus orientation over the other mainly in superficial and deep layers. Furthermore, for preferred or non preferred orientation a relative increase for the respective other α sub-band as bind observed. Hence, we report for the first time, evidence for α oscillations that mediate either attention or feature related processes with a differential layer and frequency activation profile which - combined with findings on γ - would be in line with predictions by recent theoretical frameworks [91]. Feature unspecific attention related α [251, 272]) as well as feature specific α with respect to the ongoing feature processing [91, 309] have been predicted. Since those effects could be observed not only in voxel with strong stimulus driven activation but furthermore in voxel with stimulus dependent *deactivation* we would like to emphasize the importance of studying negative deflections of the BOLD signal to explain cortical processes in full detail.

Methods and Materials

Subjects Data of 52 right-handed participants (34 ♀) between 18 and 35 years old ($\mu = 24.0$, $\sigma = 4.0$) was recorded. Subjects requiring eye sight correction were excluded due to impracticalities during the recording. Furthermore, participants were excluded if they exposed a neurological or psychiatric history or underwent neurosurgery. Informed consent was collected and participants were rewarded monetarily. The experiment was approved by the local ethics committee.

Data Acquisition Functional and anatomical magnetic resonance imaging (fMRI) data was collected using a Siemens MAGNETOM Prismafit 3T MRI scanner, employing a 64 channel whole head and neck coil. Before moving into the scanner, the subject was instructed and a short practice block of the main experiment was performed. After that the subject was placed inside the scanner. First, a high resolution (0.8 mm isotropic voxel size) whole brain volume was recorded using a T1 weighted MP RAGE sequence [310]. Repetition time (repetition time (TR)) was set to 2.2 s and echo time (echo time (TE)) to 2.64 ms. In total 224 sagittal slices of $256 \times 256 \times 0.8$ mm were acquired. Functional data was recorded using a 3D gradient-echo planar imaging (EPI) [311] partial brain protocol. In total 44 tilted coronal slices of $208.8 \times 208.8 \times 0.9$ mm from the occipital and parts of the parietal lobe, were collected per volume, including primary visual areas. The phase encoding flip angle was set to 20° at an (almost) isotropic voxel size of $0.9052 \times 0.9052 \times 0.9$ mm (TR : 3.3 s; TE: 34 ms). Slices were collected interleaved in ascending order. This protocol was used for both the main experiment and the retinotopic field mapping. See Figure 10 (B) for a visual representation of the recording protocol.

Simultaneously, **electroencephalography (EEG)** data was recorded using a 64 channel MR compatible EEG system [312] at a sampling rate of 5 kHz. Impedances were kept below 20 k Ω during subject preparation. Electrode positions were recorded using a photogrammetry based approach as described in Clausner et al. (2017) [313]. A 3D model, computed from ≈ 50 photographs of participants wearing an EEG cap, is aligned via facial features to a 3D representation of the anatomical MRI. Electrode positions are determined from the photogrammetry based 3D model, transformed into MRI space.

Additionally, eye tracking data was recorded simultaneously using an EyeLink 1000+ [314] at a sampling rate of 1 kHz (to check for proper stimulus fixation). Calibration was performed using a 3×3 grid filling the central square of the screen. For each grid point, the calibration error was kept below 0.5° visual angle deviation from the center. Note, that due to difficulties during data acquisition and insufficient data quality, eye tracking data was omitted from the analysis protocol afterwards.

The full experimental protocol included (in chronological order) the high resolution anatomical scan (8 *min*), four blocks of simultaneous EEG-fMRI recording for the main experiment (4 × 14 *min*), three blocks of population receptive field (pRF) mapping (3 × 7 *min*) using the same fMRI recording sequence, but omitting EEG data recording, and 20 resting state volumes of that sequence, but with inverted flip angle (1 *min*) for estimating field distortion (omitted from the analysis protocol). The duration of the entire experiment summed to about 150 *min*, including preparation time, a 5 – 10 *min* break between the two main experimental parts and 15 *min* after the experiment has ended for participants to wash and dry their hair.

Stimulus presentation Stimuli were projected onto a screen behind the subject’s head using an EIKI LC XL100 projector (<https://www.eiki.com/>) at a resolution of 1024 × 768 *px* and a maximum brightness of 5,000 ANSI - lumen, with 1000 : 1 contrast ratio. Light from the projector screen was re-routed using a mirror system mounted to the MR head coil to reach the subject’s field of view. The effective field of view comprised 24 × 18° visual angle at a distance of 855 *mm* relative to the subject’s eyes. Throughout the entire experiment, stimuli were presented in an otherwise dark scanner room. During the anatomical scan, participants could read the experiment instructions and remained otherwise still with their eyes opened or closed.

Main experiment Subjects performed a visual attention task, using central stimulus presentation. Left (counterclockwise) or right (clockwise) oriented gratings ($\pm 45^\circ$ relative to the vertical axis of the screen) could expose a subtle wavy pattern which served as oddball stimuli that participants were asked to respond to. The non-oddball : oddball ratio was set to 5 : 1. Stimuli were presented using "Presentation" [315] on a gray background of 50% luminance. A fixation indicator was created based on the findings of Thaler et al. (2013) [316], consisting of a black, filled circle overlaid with a white cross ("Greek cross"), housing a central fixation dot . It was found that such a stimulus yielded the highest fixation performance compared to more common fixation stimuli like simple crosses or dots. In our experiment, the central fixation dot could either be red or green indicating the subject whether to avoid blinking (red = avoid blinking).

Feature specific stimuli (left or right oriented) were constructed as Tukey-filtered gratings of 8° visual angle in diameter and a spatial frequency of 3.125 cycles per 1° = 25 cycles, presented at the central screen location. The contrast between bright and dark components was set to 70% luminance change. An area of 0.8° visual angle in diameter was cut out centrally to house the fixation mark. Stimulus gratings could be presented in either left or right orientation, deviating $\pm 45^\circ$ from the vertical axis. Additionally, oddball trials were

constructed similarly, but exposed a slightly wavy pattern of an amplitude of 0.3571° visual angle and a frequency of 0.6526 cycles per degree visual angle (≈ 4 cycles across the diameter of the stimulus area). Furthermore four different phase offsets $[0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}]$ were used in a pseudo-randomized manner. The outer edge of the stimulus, as well as the edge towards the inner cut-out where the fixation was placed, was filtered using a Gaussian kernel, to avoid sharp edges.

All settings concerning stimulus appearance were piloted to obtain a satisfactory trade-off between difficulty and accuracy. An example for an oddball stimulus can be found in Figure 10 (A).

Receptive field mapping To localize regions of interest (ROIs) of primary visual areas, a population receptive field mapping was performed in order to obtain structural locations of V1, V2 and V3 (V4 where possible) from functional data for both hemispheres [317]. One hundred twenty-eight volumes were recorded for each of the three experimental blocks. Stimulus presentation was implemented in PsychToolbox [318] using the VistaDisp software package [319]. A sequence of full contrast "checkered" bars was presented, moving from $W \rightarrow E$, $SE \rightarrow NW$, $N \rightarrow S$, $SW \rightarrow NE$, and reversed directions, in that particular order. A circular area of 18° visual angle in diameter was covered in front of an otherwise empty screen of 50% luminance. Each bar was 2.25° visual angle wide and up to 18° visual angle long (filling the central circular area 18°). The overlap with neighboring bar locations was 1.125° (half a bar's width). Sixteen different locations along the directional axis for each moving direction were sampled ($\sum = 128$ trials per block, sampling each location twice). Within each location an alternating full contrast black and white pattern was presented five times for 0.66 *ms* per cycle (= 0.66 *ms* for 2 alternations). For one of those sets of five consecutive pattern repetition cycles, one volume was recorded. For each diagonal moving direction, the pattern disappeared for the last eight locations (40 cycles) of that direction in order to allow for the BOLD response to fall back to baseline. This procedure is further described in Alvarez et al. [320]. Central fixation during pRF mapping was ensured by a fixation dot that would randomly change color between red and green on an average rate of ≈ 0.3 Hz. Subjects were asked to indicate such a color change by a button press with the left index finger. All three experimental blocks were recorded consecutively without a break. For subject *S11* only the first two blocks were recorded, since the subject had to leave the scanner because of a feeling of discomfort.

Experimental Procedure After participants were instructed and informed consent was collected, the EEG cap was fitted and electrode positions were recorded. Hereafter electrode housings were filled with an electrically conducting gel to bridge the gap between the electrodes and the skull. Afterwards the subject was placed in the scanner. Foam and pillows

helped to keep the subject’s head stable and to remain comfortable throughout the experiment. A strap of tape across the forehead provided tactile feedback of any head motion and thus supported the required avoidance of head motion. The eye tracking device was set up after that and was calibrated using a 3×3 calibration grid filling the central square at the screen. It was ensured that the deviation to the center of each calibration point was below 0.5° visual angle.

Before the main experiment started, a high resolution T1 weighted anatomical scan was performed using an isotropic voxel size of 0.8 mm. During this the subject performed a practice block for the actual task that followed hereafter. The practice block was a slightly modified version of the main task, such that the inter-stimulus-interval (inter stimulus interval (ISI)) was shortened and the ratio of oddball over non-oddball trials was increased to facilitate the training effect.

The main experiment consisted of four blocks of 60 trials each, ten of which were oddball trials that were excluded from the later analysis. Subjects were instructed to respond as fast as possible to the occurrence of such a trial by pressing the response button with their left index finger. Each experimental block started with six dummy volumes to allow for the magnetization to reach a steady state, but only the last three dummy volumes were actually recorded.

A trial was defined by the following sequence of events: 1200 *ms* prior to the onset of the stimulus, the blink indicator of the central fixation would turn to red until 1600 *ms* after the onset of the stimulus ($\sum = 2800ms$), indicating the subject to stop blinking. Hereafter the stimulus was presented, such that with a probability of $p = 0.167$ an oddball trial would occur. Each of the stimuli was presented for 1600 *ms*. If no oddball was shown and the subject did not respond by a button press, the red fixation was presented for additional 600 *ms* before turning back to green, which would end the trial. In all other cases the subject would receive feedback in form of a centrally presented text indicating hit, miss or false alarm, followed by the green central fixation. Hence, a trial would always last $1.2 + 2.2 = 3.4$ *s* of which the last 3 *s* went into the EEG analysis. During that last 3 *s* before the trial ended, MRI gradients and RF pulses were switched off, such that no MR data could be collected. This was done to ensure sufficient EEG data quality. After each trial, three partial brain 3D EPI volumes ($TR = 3.3$ *s*) were recorded, sampling the BOLD response for a single stimulus presentation of that length [321]. For a full overview of how a trial, as well as how the gaped sequence was constructed see Figure 10 (A).

The described procedure was repeated 60 times within each of the four experimental blocks, summing up to 240 trials in total per subject. Since trials could be constructed as left or right oriented, oddball or non-oddball, four possible trial types could occur (ignoring

the phase offset of the oddball trials). Since the ratio between non-oddball and oddball trials was fixed at 5 : 1, trial types were unevenly distributed between blocks but counterbalanced across blocks. Each block contained 25 non-oddball and five oddball trials for each respective orientation. Hence, each subject was exposed to 100 non-oddball and 20 oddball trials for left and right oriented gratings respectively ($\sum = 240$) across the entire experiment. In total 180 functional volumes were collected per block. This corresponds to an experimental duration of ≈ 55 min for all four blocks.

After the main experiment, participants could voluntarily rest for some minutes before the population receptive field (pRF) mapping [32] was performed. Each of the three pRF mapping blocks consisted of 128 continuous "trials" lasting for 3.3 s. To ensure central fixation, a colored fixation dot (red or green) was presented at the center of the screen. Subjects were asked to indicate a color change that occurred randomly at an average change rate of 3.3 s. Each trial corresponds to 1 TR (3.3 s), sampling each bar location twice. Population receptive field mapping required additional 21 *min* of experimental time.

Finally, 20 volumes of the very same sequence used during the pRF mapping were collected with an inverted phase encoding direction (200° flip angle). This was done for later field distortion correction, but was eventually omitted after a different analysis strategy was chosen (see below). Furthermore, a whole brain proton density scan and a whole brain inverted proton density scan was recorded, but not used in the later analysis. The total experimental duration sums up to ≈ 40 *min* of preparation time per subject, ≈ 95 *min* inside the MRI scanner and another 15 *min* after the experiment for the participants to clean their hair.

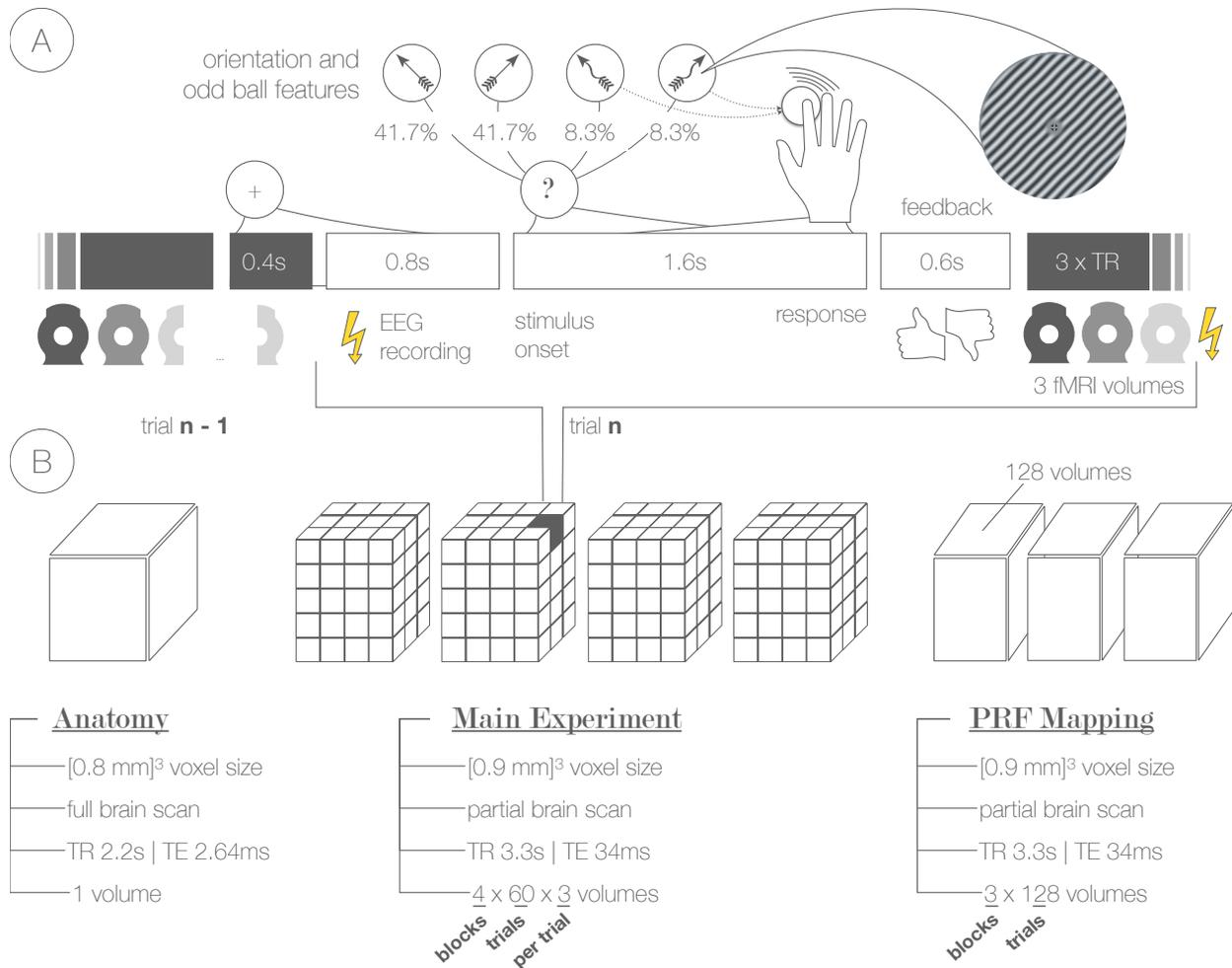


Figure 10: A) Experimental Procedure of the main experiment. 1.2 s prior to the stimulus onset the fixation indicator would turn from green to red, indicating the subject to avoid blinking. After a period of 400 ms (800 ms before stimulus onset) the last fMRI volume of the previous trial finished recording. For a period of 3 s no fMRI data is collected in order to avoid gradient artefacts in the EEG data. The stimulus presented is a left or right ($\pm 45^\circ$ from vertical axis) oriented grating. In 16.7% of stimuli the linear pattern of the grating would expose a slight wiggly pattern (see example grating). Subjects were ask to respond to those oddball trials with a button press. If no oddball trial was presented, the stimulus would remain on the screen for 1.6 s, followed by a 600 ms period where only the fixation indicator is shown. In case of a response, corresponding feedback ("correct", "false alarm" or "miss") is displayed instead. Afterwards, the fixation indicator turns back to green, indicating the subject that the period to avoid blinking has ended. Now, three consecutive 3D EPI volumes (TR : 3.3 s) are recorded, before the next trial starts. Due to the gapped sequence, a significant signal drop-off over the course of three consecutive volumes can be observed, which is illustrated by the color of the MRI machine pictograms. **B) Experimental Protocol.** Each session would start with the recording of the anatomical MRI, during which the participants practiced the task. Afterwards four experimental blocks of the main experiment were recorded. The exact sequence of events for each trial is depicted in A). For each of the four blocks, 60 trials have been recorded. For each trial 3 s of gradient artifact free EEG data and three 3D EPI fMRI volumes were recorded. The main experiment was followed by three blocks of a pRF mapping experiment, where each block contained 128 trials.

Data processing Data analyses were performed using the following software packages and toolboxes: analyzePRF [32, 322], ANTs [323], FieldTrip [324], Freesurfer [325], FSL [326], janus3D [313], Metashape [327], MRICron [328], MRI Volume Masker 3000 TM [329], MrVista [319], OpenFmriAnalysis [330], SPM12 [331], Workbench [332], including respective dependencies in either Bash, Python or MATLAB. All analysis scripts can be downloaded from <https://github.com/TommyClausner/laminarfMRIv2>.

fMRI Motion Correction and Co-registration Motion parameter estimation and correction was done using ANTs. As a first step, each set of volumes (three blocks pRF mapping; four blocks main experiment), was striped by the first three initial volumes that were collected as dummy volumes to allow for the BOLD signal to reach a steady state. Afterwards a manually drawn brain mask was created for every first volume of the first main experiment block and the first retinotopy block using MRI Volume Masker 3000 TM. Automatically generated masks were manually “fine tuned”, such that the outer boundary was enclosing the gray matter as close as possible. Extensive parts of cerebrospinal fluid, fatty components, arteries and other tissue were carefully excluded from the masks. However, the level of detail on which this operation was performed was kept above single voxel selection. Resulting masks were used to constrain motion parameter estimation and to correct the anatomical segmentation performed by Freesurfer within the respective region of interest (the field of view of the functional scan). The actual motion parameter estimation was then performed in two stages, but with similar parameter settings. In the first stage all volumes of one recording block were registered to the within-average over time of that block. During the second stage all newly computed within-block averages were registered to the first volume of the first block of the main experiment. Thus all blocks, including pRF mapping, used the first volume of the first block of the main experiment as the final reference. For both stages, the actual computation was performed by `antsRegistrationSyN.sh` that was provided by ANTs using the default settings that were adjusted when necessary. While for the first stage a rigid body transformation was used, an affine transformation was computed for the second stage. The initial linear transformation in that stage was followed by a non-linear transformation using symmetric normalization (SyN) [333].

A similar approach was used for functional to anatomical partial volume co-registration. Again, the default settings of `antsRegistrationSyN.sh` were used for that purpose. Hence, a 3 stage approach, consisting of a rigid, an affine and a non-linear SyN-registration, was chosen. However, for co-registration the T1 weighted image was registered to the functional data. This was done, because the later constructed laminar profile needs to match the functional data space as close as possible and hence slight distortions due to the spatial transformation was prevented. All estimated motion parameters were combined and applied in a single operation to ensure that functional data was interpolated only once [334].

Anatomical Segmentation Since the ANTs co-registration procedure yields forward and backward transformation matrices and non-linear transformation volumes, the manually drawn functional masks could be registered to native T1 space. This was done in order to ensure proper anatomical segmentation, of the FreeSurfer function `recon -all` particularly in the area covered by the functional scans. Precisely, `recon -all` was called, respective brain masks obtained as described above and parietal parts of the full brain mask (the area covered by the functional data), were replaced by manually drawn masks of the functional data. Hereafter the estimation of pial and white matter surface boundaries was recomputed using `recon -all` but including the function argument `-autorecon-pial`. Corrected pial surfaces and uncorrected white matter surfaces were used as boundaries for the laminar segmentation as this procedure makes use of the surface’s boundaries. White matter boundaries were kept untouched for that matter.

Population Receptive Field Mapping Population receptive field (pRF) mapping was performed as implemented in the open source tool-box `analyzePRF`. More detailed information about the algorithmic implementation can be found in the reference literature [32, 322]. Binarized versions of each stimulation frame served as spatial regressors for the underlying general linear model (GLM). Each of the presented 64 unique bar locations (including blanks) were thresholded, such that the background received a value of 0 and the entire bar irrespective of the checkerboard pattern received a value of 1. In order to save computation time, stimuli were downsampled from screen resolution to a resolution of 192×192 *px*. A Savitzky–Golay filter with a filter window of 61 TR s (201 seconds) was applied to the data. The data then was converted into percent signal change relative to the median. Whereas the Savitzky–Golay filter was applied for each experimental block separately, percent signal change was computed over all blocks combined. Based on a GLM - including third order polynomials - parameters were estimated for orientation (angle), distance to the center of the screen (eccentricity) and the explained variance per voxel (R^2). The gray matter mask, obtained from the anatomical segmentation, was applied and only gray matter voxel locations were fed into the pRF analysis. Based on those maps, regions of interest (V1-4) were manually labelled using Freeview. To facilitate the manual drawing process, a functional atlas [335] containing all regions of interest was anatomically fitted to the functional data beforehand. Fitted regions from the atlas were overlaid together with the results of the pRF mapping onto the inflated pial surface as obtained from FreeSurfer. Marked labels were then transformed into volumetric data and into functional data space using previously computed co - registration transformation matrices and volumes.

Estimation of Cortical Layers

Laminar segmentation was performed using co - registered gray and white matter boundaries as references for upper and lower bounds of the segmentation. In order to resolve

cortical depth precisely, the curvature of the anatomical boundaries was taken into account. This is necessary since the relative thickness of cortical layers varies depending on the cortical curvature [336]. Each voxel covered by the gray matter mask, received a weight as a function of its volume belonging to each of the shell-like meshes forming the boundaries. If a layer boundary would cut the voxel exactly in half, adjacent layers would receive a weight of 0.5 each. Hence, voxels were not separately treated as belonging to different layers, but rather their signal was seen as a weighted mixture coming from different layers. See Figure 11 for a visualization. Thus, a voxel located towards the white matter boundary would contribute more to the signal generated in deeper layers - receiving a higher weight - as compared to a voxel being closer to the surface, which would receive a lower weight at the reference location [34]. Layer weights were computed using the open source toolbox OpenFmriAnalysis. As a result, five layers were obtained (CSF, superficial, middle, deep and white matter layer). Layer profiles were combined with all regions of interest obtained from the pRF analysis, for both hemispheres separately.

EEG data processing The major goal of the EEG data preprocessing was to optimize noise suppression for each frequency band of interest (α , γ). Since linearly constrained minimum variance (LCMV) beamforming works such that in order to solve the inverse problem, one equivalent current dipole location will be used to fit the sensor data, while all surrounding signal sources will be suppressed [337], unsupervised noise reduction can be performed. Note, that in previously published literature a supervised signal decomposition was used based on ICA [238]. This approach requires the manual selection of target components for each frequency band. By removing all non-target components, noise can be suppressed and the resulting signal will only contain the data of interest. However, beamforming has the major advantage of being able to perform unsupervised noise suppression [338]. During the piloting phase, beamformer methods have successfully been applied to EEG data collected in an (f)MRI environment and yielded accurate source reconstruction results. This has also been demonstrated before [339, 340]. Since the main hypotheses of the presented study do not directly address the frequency responses themselves, but rather their relation to the BOLD response, a subjective selection of EEG data that fits this purpose, still acts as valid scientific strategy and does not result in a case of "double dipping".

Data was selected for both non-oddball conditions (= 25 trials per orientation per block = 200 trials per subject). Trials were only excluded in case of a false alarm response, which was the case in 2% of all trials. Low (α) and high (γ) frequency bands were processed separately to extract the desired response patterns. The data was filtered for the lower frequencies using a pass band between 2 and 32 Hz and for the high frequencies between 20 and 120 Hz respectively. A 50 Hz dft filter to suppress power line noise was applied to the latter as well. Afterwards the noise covariance matrix was estimated for a baseline win-

dow of -500 to -100 *ms* relative to stimulus onset for every trial separately. EEG electrode locations were obtained from the photogrammetry based 3D model and co-registered using the face shape to the anatomical MRI using janus3D, as described in Clausner et al. (2017) [313]. A finite element model (FEM) was computed from the high resolution anatomical T1 based on the FieldTrip-SimBio pipeline [341]. FEM models expose increased localization performance as compared to boundary element models due to their increased spatial resolution with respect to varying tissue types [342, 343]. The leadfield was computed from the EEG electrode positions and the FEM model. Sources were modeled as equivalent current dipoles at locations limited to the respective coordinates of voxel included in the gray matter mask. Dipole orientations were derived from the cortical curvature and thickness, since this is crucial for a precise mapping especially in EEG [344]. Workbench was used to compute the surface normals that connect pial and white matter surfaces. The orientation of the resulting vectors then served as the dipole orientation for each respective location. The described procedure was done separately for the left and right hemisphere in order to obtain separate filter weights, since distinct source activity for both was to be expected [345]. LCMV beamformer source estimation was performed using a noise regularization of $\lambda = 0.1$, where 0.1 refers to 10% of the average amplitude of the noise covariance matrix. The resulting weight matrices were applied to the band pass filtered data in order to obtain virtual channels at the corresponding equivalent current dipole locations. A spectral analysis was performed on each virtual channel separately using a multi-taper approach and the Slepian sequence [346] also known as discrete prolate spheroidal sequence (DPSS). Seven tapers were used for high and three for low frequencies. The frequency domain was smoothed using either a kernel width of ± 10 Hz for high or ± 2.5 Hz for low frequencies. Afterwards the virtual channel with the highest average amplitude change between $8 - 12$ Hz (α) and $50 - 70$ Hz (γ) relative to baseline, was selected. As a baseline period, a time window from -0.3 to -0.1 s for high and at -0.3 for low frequencies was chosen. Since gradient artifacts caused by ringing of the gradient coils of the MR machine could be observed prior to -300 *ms* relative to stimulus onset, this time stamp served as lower bound. For low frequencies only a single time point at -300 *ms* was chosen, because a pre-stimulus α decrease was expected. pre-stimulus α has often been related to visual detection performance [347] and might thus reflect different processes than the actual target time interval. In fact a small but visible α decrease could be observed 250 *ms* prior to stimulus onset. The data was transformed into the \log_{10} ratio between the average baseline and each sample point in a time window between 0.1 and 1.6 s after stimulus onset for each hemisphere. The virtual channels with the time-frequency transform yielding the highest average response was chosen to be the "best" channels that were later used to build the regressors for combined EEG-fMRI analyses. In total four dipole locations (i.e. the time-frequency transform of corresponding virtual channels) were selected for each subject individually: Two for each hemisphere and two for the separate frequency bands. All steps previously mentioned were implemented in MATLAB R2021a [348] using the open source toolbox FieldTrip. See Figure 15 (B) for a depiction of the average response.

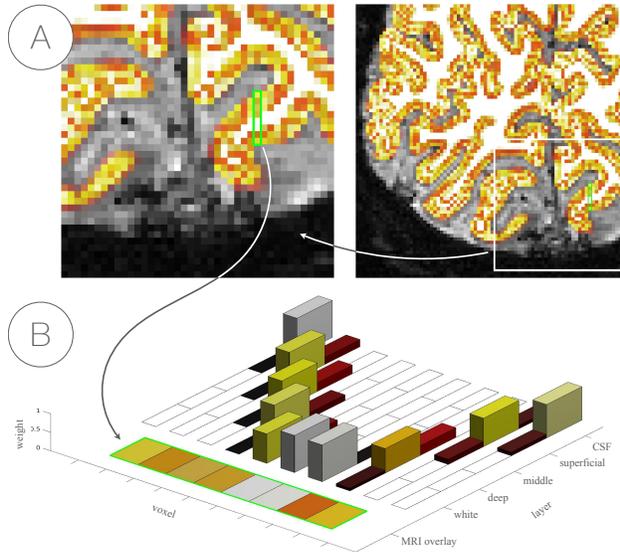


Figure 11: Laminar weight profiles: **A) functional MRI overlaid with layer profile maps.** A slice and a zoomed in section of one volume of one subject of the fMRI data, overlaid with probability maps as obtained from OpenFmriAnalysis [330]. Shell like meshes have been computed, separating five distinct layers: cerebrospinal fluid (CSF), superficial, middle, deep and white matter. Layer weights are computed as the fraction of each voxel that lies between two shells. Weights for all layers but CSF have been used as overlay for the functional data to obtain the respective image. **B) Weight profile for a set of voxels.** For a selected amount of voxel (green box), the weight profile over each layer is depicted. Across layers, values underlying the bar plots sum up to 1. In the later combined EEG-fMRI analysis, correlations between EEG signal and BOLD response for each voxel are multiplied with the respective layer weights.

sen in accordance with previous literature [238]. Furthermore the average white matter signal and the average residual signal (signal after regressing out gray and white matter signals) were included. Due to the way the data was recorded (gapped sequence), a significant signal drop-off could be observed for every three consecutive volumes which was included as a nuisance regressor as well. All motion parameters (translation along and rotation around x, y, z) and their first derivatives were included as well as a set of high pass filters modelled as five sines and five cosines. Those five sine and cosine waves were constructed, such that they would span one to five full cycles across one experimental block.

Combined EEG-fMRI analysis The general logic of fitting EEG time-frequency regressors individually for each bin to the BOLD signal for different cortical regions and across layers, follows coarsely what is described in Scheeringa et al. (2016) [238]. Several steps have been undertaken to prepare the EEG and fMRI data for the later combined analysis and will be described in the following.

Task regressors were built for all trials that did not contain: oddballs, strong EEG data artifacts (e.g. jumps, muscle, etc) or false alarm responses. Regressors were built separately for left and right oriented gratings. Nuisance regressors contained all trial and response combinations that were not included in the task regressors (e.g. false alarm trials). Additional regressors contained blink or artifact trials, button presses and reaction times. All the aforementioned regressors were convolved with the hemodynamic response function as built into SPM12. Reaction time regressors were treated as parameter modulators. Thereby the onset of the modulation was set to the average reaction time for each individual block and the actual reaction time as the modulator value. This procedure was chosen

EEG data regressors were built on time-frequency resolved virtual channel data, obtained as described above. This was done by convolving different parts of the EEG data with the hemodynamic response function that comes with SPM12. A regressor was built for each frequency and time bin separately. Due to the way time-frequency data was obtained, one frequency bin span 0.5 Hz for low or 2.5 Hz for high frequency data (a single data point) and time bins were set to 400 ms length. Time bins were shifted by 100 ms intervals. Thus consecutive time bins contained overlapping data. The data within a respective time frequency bin was averaged and treated as parameter modulators for the hemodynamic response function. Thereby the onset was set to the mid data point of the time domain and shifted by one to three TR s to match the corresponding three volumes that were recorded after each gap where EEG data was recorded. This approach yields a multitude of EEG data regressors, where a separate model for each time frequency bin was constructed. Task and nuisance fMRI regressors were kept fixed for each model. Thereby three kinds of responses were taken into account: all trials ("activation") and trials where the stimulus was either left or right oriented.

Before the combined EEG-fMRI analysis could be conducted, a first level fMRI analysis on the motion corrected data was performed in order to identify voxel of interest. Thereby two contrasts were used: Both stimulus orientations vs. baseline (activation: **A**) and the contrast between both stimulus orientations, that is left - right (contrast: **C**). While the first was used for additional analyses that are explained later, the second served as the contrast on which most of the following experimental analyses were based on. In general, first level analyses have been conducted for each subject separately, but on all experimental blocks combined. The data has been z-scored beforehand only along the time domain, separately for each voxel and experimental block. Using aforementioned nuisance regressors as control parameters, the result of the first level analysis was a separate activation t-map for **A** and **C**. For the later inferential analysis, t-value thresholds needed to be defined in order to determine an appropriate voxel sub-selection. Thereby, a higher threshold indicates higher specificity for the respective stimulus and a lower threshold increases the signal-to-noise ratio (SNR) by including more data. This changes the number of voxel selected. In previous publications, employing a similar experimental setup, 500 voxel with highest feature specific activation [349, 350] or the top 10% activated voxel [238] were selected. A study by Markuerkiaga et al., specifically designed to assess the number of voxel required, finds 250 voxel (for 3TfMRI) to yield the best contrast-to-noise ratio (contrast-to-noise ratio (CNR)) [351]. However, since all previously mentioned publications set a more or less arbitrary threshold and the last did not take the correlation with EEG into account and potential changes in SNR and CNR for those cases, three thresholds have been selected to eliminate eventual uncertainties: 5%, 10% and 25% of most specific (or activated) voxel. The respective average (standard deviation (SD)) number of voxel for each threshold and first level contrast can be found in Table 1 for each of the specified ROIs. Since voxel were selected based on the t-value before

the gray matter and ROI mask was applied, the respective selections do not necessarily imply a precise five-fold increase in the respective numbers.

top # voxel	\mathbf{A}^+			\mathbf{A}^-		
	V1	V2	V3	V1	V2	V3
5% μ (σ)	327 (223)	443 (276)	432 (268)	235 (163)	326 (205)	288 (200)
10% μ (σ)	694 (358)	926 (436)	880 (407)	554 (271)	745 (337)	660 (333)
25% μ (σ)	1727 (608)	2228 (749)	2050 (666)	1515 (505)	1965 (629)	1734 (575)
	\mathbf{C}_L^+			\mathbf{C}_R^-		
5% μ (σ)	54 (32)	68 (38)	61 (28)	49 (31)	64 (30)	60 (30)
10% μ (σ)	273 (85)	349 (106)	314 (97)	256 (92)	332 (112)	307 (105)
25% μ (σ)	1354 (349)	1719 (432)	1524 (362)	1291 (370)	1650 (450)	1500 (384)

Table 1: **Number of voxel per condition.** For each ROI the average number of voxel over participants and standard deviation (in parentheses) is shown, separately for each selection threshold. These thresholds reflect the highest 5%, 10% and 25% t-values for the respective condition and ROI (lowest t-values for the right column). Since gray matter and ROI masks were applied after the respective voxel selection, numbers do not necessarily increase five-fold. See Figure 15 (D) for a visualization of the number of voxel in V1 with respect to different condition sub-selections.

General activation General BOLD activation patterns for stimulus compared to baseline that have been hypothesized to be linked to attention [354] (see Figure 15 (C) left) have been related to EEG time-frequency power changes. Each time-frequency bin specific EEG regressor for left *or* right orientation combined with selected voxel of \mathbf{A} for each of the specified thresholds. Since the respective first level analysis of the fMRI data to determine activation compared to baseline results in a t value map for each subject with positive *and* negative values within respective ROIs. Since spatial attention has been demonstrated to cause increased BOLD activity over primary visual regions in retinotopically corresponding attended areas and a relative decrease in unattended regions [354] and has been linked to a general modulation of a multitude of attention related brain networks [355, 356], it is assumed that \mathbf{A} reflects some similar mechanism. The data has been split between positive (\mathbf{A}^+) and negative \mathbf{A}^- generally active voxel at $t = 0$. In addition to that, the signal of the 25% most feature selective voxel for one orientation compared to the other ($\mathbf{Pc} - \mathbf{nPc}$, see below) has been used as a control regressor for the \mathbf{A}^- selection. Since, voxel that expose a positive t value compared to baseline in one of the contrasts are thought to most likely include feature *and* attention related signal components, this analysis has been conducted for \mathbf{A}^- only (see also Figure 13 and Figure 14 top). A separate model was built for each time-frequency bin specific EEG regressor, including fMRI task and nuisance regressors and separately for each voxel (see Figure 12 (A)). This procedure results in a β weight for each voxel that was multiplied with the corresponding layer weights. Afterwards, the result was averaged over voxels yielding a *layer* \times *frequency* \times *time* matrix for \mathbf{Pc} and \mathbf{nPc} for every ROI, threshold and subject (see Figure 12 (C)). All main analyses were performed for a time window ranging from 0.1 s to 0.8 s after stimulus onset, for which corresponding time bins were averaged. This was done because the average reaction time to odd-ball stimuli was 759 ms. Since the main stimulus processing is assumed to take place beforehand, but in order to include as many data points as possible, the time bin including 0.8 s after stimulus onset (50 ms after the average reaction time) was included as well. See Figure 12 (A-C) for a visualization.

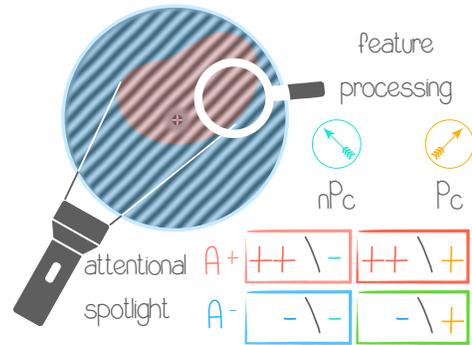


Figure 13: **Attention and feature related processes.** Both, attention and feature related processes modulate the BOLD signal. Since the relative decrease for the not preferred feature as compared to the preferred feature is expected to be lower than the general modulation by (spatial) attention. A differentiation between features could be obtained by considering generally *deactivated* (unattended) voxel, as the respective feature contrast is expected to be less influenced by global changes. In addition to that, feature specific signals can be used as control signal for the correlation with deactivated voxel to obtain a general deactivation measure. It has been hypothesized that both processes (spatial attention and feature processing) are implemented in separate mechanisms in the α band [91].

Contrasting \mathbf{Pc} and \mathbf{nPc} For the main feature contrast analyses, each time-frequency bin specific EEG regressor for left and right orientation was combined with selected voxel of \mathbf{C} for each of the specified thresholds (see Figure 12 (B)). Since \mathbf{C} was constructed, such that positive t-values would indicate a voxel’s preference to the left orientation (\mathbf{C}_L^+), whereas negative t-values would indicate a preference for right oriented stimuli (\mathbf{C}_R^-), four possible combinations arise. Those can be condensed into a preferred ($\mathbf{Pc} = \mathbf{P}$ referred based on **contrast**) and a non-preferred ($\mathbf{nPc} = \mathbf{n}$ ot \mathbf{P} referred based on **contrast**) condition. Within the preferred condition the EEG regressor specific to left stimuli would be combined with \mathbf{C}_L^+ and the EEG regressor specific to right oriented stimuli with \mathbf{C}_R^- and accordingly for the non-preferred condition. Thus the main feature contrast could be described as $\mathbf{Pc} - \mathbf{nPc}$. Note, that this analysis has been conducted irrespective of \mathbf{A}^\pm . Other regressors and model parameters have been set exactly as for the general activation (see above). The contrast between \mathbf{Pc} and \mathbf{nPc} yields a final result in form of a *layer* \times *frequency* \times *subject* matrix for each respective condition of interest, threshold and ROI. In order to separate feature and attention related signal components, the respective \mathbf{Pc} or \mathbf{nPc} voxel selection can be combined with the \mathbf{A}^+ and \mathbf{A}^- selection as used for the general activation (see above). Using this method, e.g. feature specific voxel can be identified that expose e.g. a general positive or negative BOLD response (see below and Figure 14 bottom).

Feature and frequency specific (de-) activation Spatial attention has been associated with a spatiotypical increases in the BOLD signal in some primary visual areas, such as MT [355] but furthermore, to receptive field-like increases over attended and decreases over unattended sites [354]. In addition to that, negative relative changes in the BOLD signal have been reported to be linked to specific attention networks [356]. Multiple α sources are expected depending on the underlying function (e.g. *spatial* attention and *feature* processing). Attention related α , potentially frontally driven [251], would differentially modulate the excitability of attended and unattended receptive field locations [91]. Such attention-related α should be found in \mathbf{PcA}^- , \mathbf{nPcA}^- and \mathbf{PcA}^- -*unspecific* voxels as attention process should be receptive field specific and decrease both specific and unspecific voxels. Feature specific α on the other hand should be specifically related to \mathbf{Pc} (and potentially inversely to \mathbf{nPc}). By comparing the relative

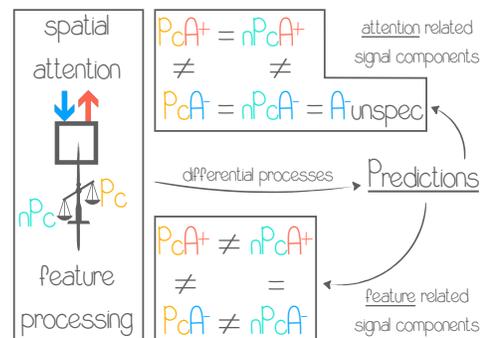


Figure 14: Attention and feature separation. Attention and feature related (α) processes, are expected to correlated differentially with the BOLD signal. General activation (attention) related processes are separated at $t_A = 0$ into activated and deactivated voxel. A difference between \mathbf{Pc} and \mathbf{nPc} within each respective condition of \mathbf{A}^\pm is expected as well. Feature specific components between \mathbf{A}^+ and \mathbf{A}^- are not expected for \mathbf{nPc} , but for \mathbf{Pc} (attending and processing and forwarding target feature information).

deactivation of the BOLD signal between feature conditions, feature selective but not attention activated neuronal responses can be related to α band oscillations. In a first step, feature selective voxel have been determined by contrasting both feature conditions in a first level analysis and extracting the respective difference t-map. From this t-map, top e.g. 5%, 10% and 25% voxel prefer feature one, whereas bottom e.g. 25%, 10% and 5% voxel prefer feature two. This selection has been performed for each ROI and combined with the sub-selection of voxel that resulted in a negative t-value for the comparison between *any* feature compared to baseline (global deactivation; \mathbf{A}^-). To further separate attention related deactivation from unattended feature specific signals, the signal of the top and bottom 25% most least activated voxel from \mathbf{C} have been averaged and used as control regressor for the combined EEG-fMRI analysis. The resulting deactivation does hence not depend on feature related processes anymore. Negative BOLD responses from mostly feature *specific* and less attention driven voxel have been correlated with α band power as well. Contrasting \mathbf{P} and \mathbf{nP} eliminates certain aspects of (spatial) attention already because it can be assumed that spatial attention is not reflected feature selectively in the same process [357]. Hence, combining \mathbf{Pc} and \mathbf{nPc} with \mathbf{A}^- enables the investigation of feature but not attention selective voxel. See Figure 15 (D) for a visualization of such a sub-selection. The width of the base-line of the figure corresponds to the number of all voxel (average over participants) in V1. From there, voxel are split roughly in half by contrasting the two features (one group preferring one and another group of voxel preferring the other condition). From those that prefer a respective orientation over the other, a separation is made between those that respond positive or negative to the respective *non* preferred stimulus orientation compared to baseline. Again, from those the still respond positive to the presentation of the not preferred orientation those that expose a smaller or larger response compared to baseline when the not preferred stimulus was presented. A similar logic is applied to the combination of $(\mathbf{n})\mathbf{Pc}$ and \mathbf{A}^\pm .

In order to investigate whether α band oscillations respond differential to different stimulus features (\mathbf{Pc} or \mathbf{nPc}), the spectrum for the correlation with \mathbf{Pc} or \mathbf{nPc} has been normalized and contrasted. Since stimulus induced α decreases could be observed for both \mathbf{Pc} and \mathbf{nPc} the respective frequency-feature selectivity is overshadowed by general activation related α effects. Hence, the frequency spectrum for both conditions has been normalized separately for each cortical layer across the α band (8 Hz to 14 Hz). This indicates the relative change or shift towards upper or lower α frequencies relative to the other. After z-transforming the data as described for each subject individually, both groups have been contrasted using a two sided cluster permutation test [352]. It is hypothesized, that the layer-frequency profile when comparing \mathbf{Pc} and \mathbf{nPc} is different for \mathbf{A}^+ compared to \mathbf{A}^- , because the combined activation of \mathbf{A}^+ and \mathbf{Pc} might reflect the ongoing joint process required to perform the task (see Figures 13 & 14).

Replication of Scheeringa et al. (2016) [238] Scheeringa et al. used a slightly different stimulus material, but methodologically comparable analysis strategies. Stimuli were composed of inwards moving concentric rings, where a speed change of the concentric movement would occur after $1400\text{ ms} \pm 200\text{ ms}$, which participants had to indicate via a button press. To match the stimulus response within the voxel selection as close as possible, only **Pc** voxel have been taken into account in the present study. Neuronal populations in primary visual areas (especially V1) respond specifically to certain stimulus orientations [358]. Using concentric rings a multitude of orientation would be activated by means of preferred orientation. Since the analysis was conducted on a trial by trial basis, **A** could not serve as the respective contrast of reference, since not the specific orientation, but rather the presentation itself was modelled. Instead **Pc** would resemble the respective stimulus specific orientation for each trial more closely. Furthermore, Scheeringa et al. included the full trial into the analysis. For this reason the full trial was included in the replication as well (0.1 s to 1.4 s in the present paper). All remaining analysis steps were undertaken exactly as described above. Since the present study can be viewed as an extension of Scheeringa et al., the replication analysis would serve as a sanity check.

Inferential statistics Due to the high dimensionality of the data, a frequency bin of interest (FOI) for low and high frequencies has been selected separately for later inferential analyses to limit the feature space for the cluster based statistical test. For low frequencies the FOI was set the frequencies between 8 Hz and 14 Hz, whereas for the high frequencies the FOI was defined as frequencies between 50 Hz and 70 Hz, which covers the peak response frequencies found in the average EEG data (see Figure 17). Within the respective range, a single tail cluster permutation test [352] has been conducted separately for each ROI, threshold and FOI for α and γ . Each cluster has been further processed by means of an auto-regressive rank order similarity (aros) test [353], which was developed by one of the authors in order to determine layer specificity. Since clusters span layer and frequency bins unevenly, the data that was used for the later aros test was collapsed over frequencies, such that the lowest and highest significant frequency bin served as the boundary over which the frequency domain was averaged. This was done irrespective of the respective cluster size within a specific layer (see Figure 12 (D)). Nevertheless, testing for layer specificity is not straight forward due to issues related to multiple comparison and non-normal distributed data. To circumvent this, Scheeringa et al. (2016) [238] tested for EEG-fMRI layer specificity by fitting layer profiles for α and γ to the respective other using an ordinary regression and tested whether β weights differed to zero. While this approach is suitable for demonstration purposes, it does not reveal the exact nature of those differences between layers. While Scheeringa et al. proved the concept of layer specific feature extraction, the present paper aims to determine the relational activity across layers depending the feature specific response as well.

The fundamental idea behind the aros test is whether group averages (i.e. averages of the signal of cortical layer in the present case), can be ranked and whether this rank order is explained significantly better by the data than it would if the data could not be meaningfully sorted (i.e. is shuffled). This is achieved by transforming the group averages into unique rank order values and computing the average fit of the data to this rank order. In a second step data points are shuffled between the groups and the same procedure is applied (i.e. computing the rank order of the mean and the average fit of the now shuffled data to the new rank order). Repeating this permutation step a large number of times yields a permutation distribution, to which the initially computed fit value of the un-shuffled data is compared. Rejecting the null hypothesis would result in the assumption that the rank order of the group averages indeed can be explained by the data significantly better than it would if the data points could not be meaningfully sorted into those groups. Thus, in the present case it could reveal how the correspondence between the EEG and fMRI signals could be sorted across layers. However, statements about the magnitude of the difference between two (or more) layers cannot be made. This approach provides insight about the specific activation profile across layers for specific conditions within a significant *depth* \times *frequency* cluster (see Figure 12 (D)).

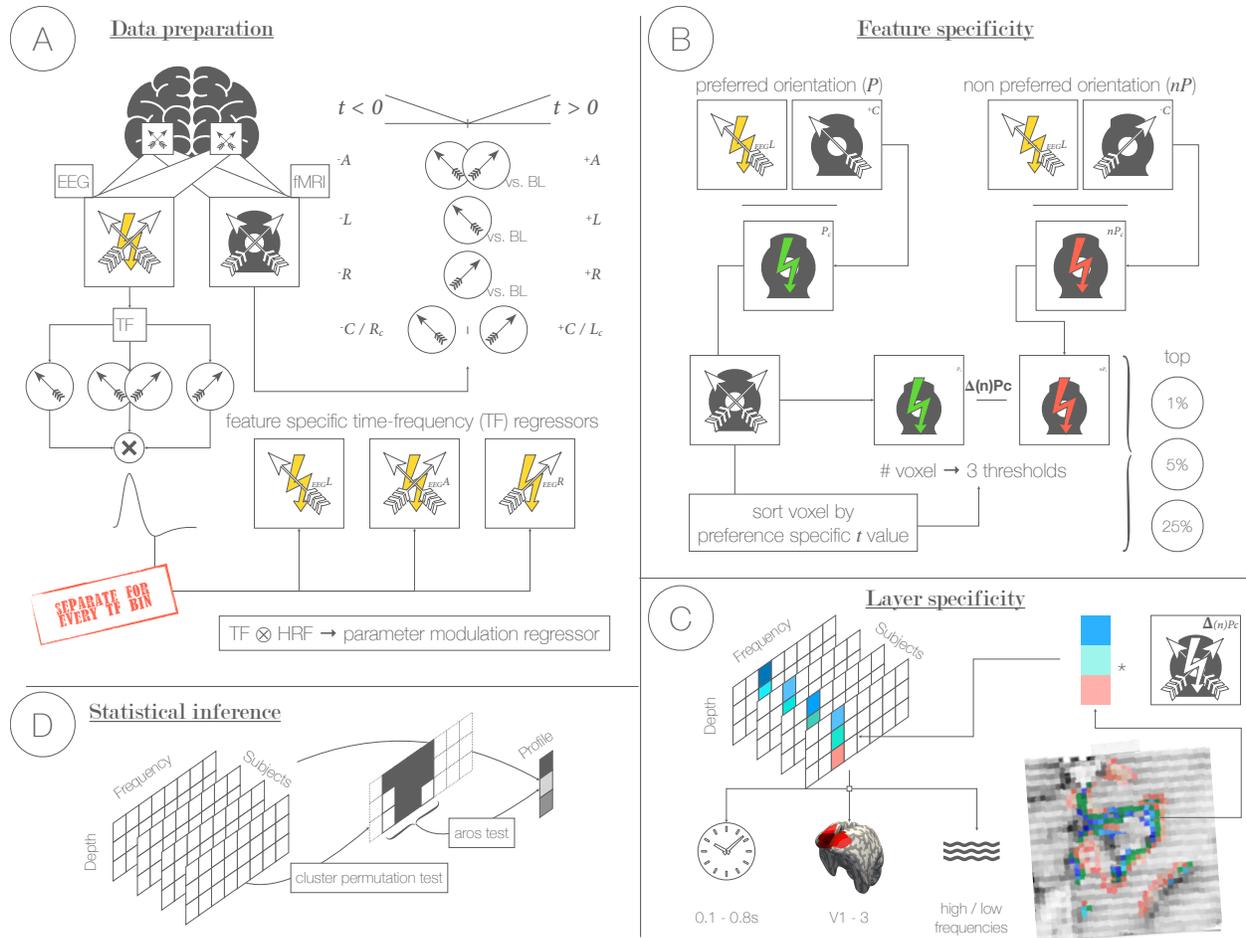


Figure 12: A) Orientation specificity. Orientation specific regressors for the EEG data were constructed based on the time-frequency transformed virtual channel data in each hemisphere that exposed the highest α decrease or γ increase respectively. Three sets of EEG derived regressors were created: specific for left or right orientated stimuli or general stimulus presentation (activation). Trial-by-trial data for each condition was modelled as parameter modulation regressor separately for each time-frequency bin. General task and nuisance regressors were used as control regressors. In order to determine voxel specific orientation preferences, a set of first level analyses were conducted. As result, t-maps for general activation (A), left and right oriented stimuli vs baseline (L and R) and the contrast between left and right (C) could be obtained. For C, orientation preference depending on the sign of each value of the t-map is indicated. **B) Main contrast ($P_c - nP_c$).** Preferred and non-preferred orientations were defined, such that the orientation specific EEG or task regressor could either match the orientation preference of the voxel derived from C (preferred: P_c) or not (non-preferred: nP_c). After obtaining β coefficients for P_c and nP_c , the main contrast was computed. Due to uncertainties about the trade-off between high specificity (restrictive t-value threshold) and high SNR (liberal t-value threshold), three different thresholds have been selected. Those included the highest 5%, 10% or 25% t-values of either P_c or nP_c . **C) Layer specificity.** Voxel specific β values are multiplied by the layer profile weights. Over all five layers (CSF, superficial, middle, deep and white matter layer), each voxel contributed a fraction of its signal to those layers, depending of the proportional volume of a voxel within each shell like mesh of each layer. Hence, β coefficients for each voxel are seen as the sum of β coefficients of the fraction a voxel is contained within a layer. This was done separately for each hemisphere and visual region and low and high frequencies. **D) Statistical inference.** Depth by frequency resolved data was tested against the hypothesis that there was no significant relationship between the EEG and fMRI data for each respective condition, using a cluster permutation test [352]. Resulting clusters were averaged over frequencies for the widest possible window. The layer profiles of the averaged clusters were tested against the hypothesis that the layer profile is as likely as any other layer profile - under the assumption of exchangeability of the data - using an auto-regressive rank order permutation (aros) test [353].

Results

Behavioral data and basic task effects Behavioral and basic task effects were only analyzed descriptively. No inferential statistics have been conducted, since those analyses mainly served as sanity checks or reflect intermediate results that were used for the main analyses. The main purpose of these analyses was to verify subject’s compliance to the task (behavioral results) and expected functional result patterns (e.g. expected EEG time-frequency responses). The shown results represent the data as it was used to conduct the final combined EEG-fMRI analyses.

On average (SD) participants responded correctly to the stimuli in 94% (8%) cases with a false alarm rate of 2% (3%) to non-oddball stimuli and a miss rate of 5% (7%), indicating the participants performed the task adequately well and complied to the instructions. The average (SD) reaction time was 759 *ms* (131 *ms*). See Figure 10 (A) for a graphical representation of the task.

The EEG signal, used to construct regressors for the final combined EEG-fMRI analyses, was obtained by a frequency analysis of each virtual channel, computed for each voxel location of the gray matter in V1. In each hemisphere, one virtual channel for low and one for high frequencies was selected that exposed the strongest α decrease and γ increase respectively. Figure 15 (B) depicts the average EEG response of the selected, time-frequency transformed virtual channel of all participants. For demonstration purposes, only low and high frequency channels of the left hemisphere are shown. By visual inspection, the main α band decrease was determined to last from 0.2 *s* to 0.8 *s* after stimulus onset and for a frequency range of 8 *Hz* to 14 *Hz*. The main γ band increase was determined between 0.1 *s* and 0.7 *s* after stimulus onset for a frequency range between 50 *Hz* and 70 *Hz*. A time window between 0.1 *s* and 0.8 *s*, was used for the main analyses, since both time windows for α and γ responses are covered. Furthermore, this time window ensures that only - but to the largest extent possible - time bins went into the analyses that were related to the actual stimulus processing ($\mu_{RT} = 759$ *ms*).

Feature specific fMRI BOLD responses, exposed a clear pattern with respect to **Pc** and **nPc**. Regression β coefficients for orientation specific regressors representing the preferred stimulus orientation for a set of voxel were higher than for non-preferred orientation specific regressors. This pattern was preserved for each of the three thresholds (5%, 10% and 25%), as well as for the average over all voxel that exposed a positive t-value within the respective contrast. Furthermore, negative t-values within the respective contrasts were more negative within the preferred than non-preferred orientation. This pattern is true for all ROIs and is exemplified for V1 by Figure 15 (A). Additionally it has been checked, whether t-maps for the

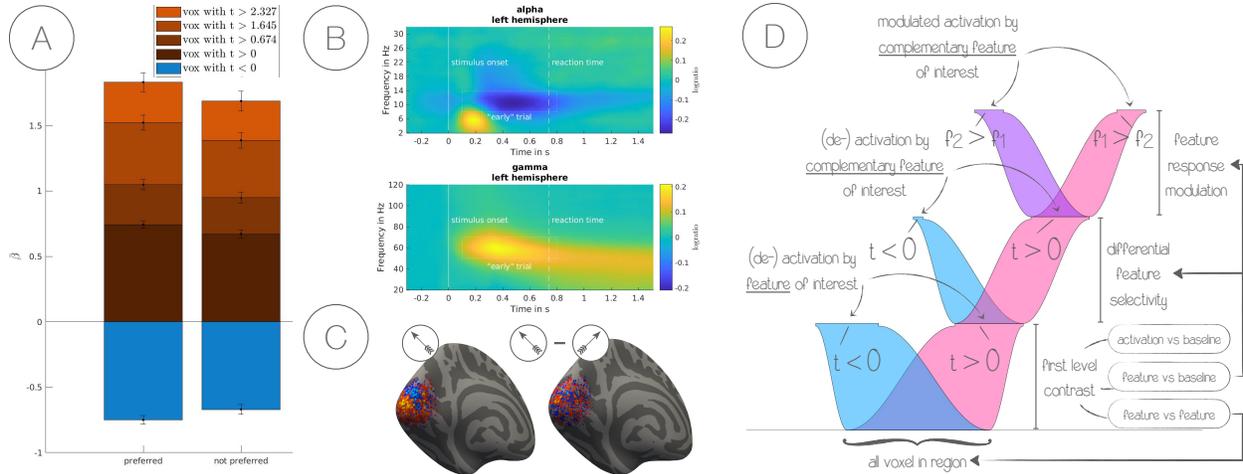


Figure 15: Task specific effects: **A) Average BOLD response to P_c and nP_c .** A general linear model was computed for the relationship between task regressors and the BOLD signal for orientation specific voxel selections. β coefficients are consistently higher for voxel that prefer the respective stimulus orientation over the non-preferred. This trend is observed when including all voxel with a positive β coefficient and pre-selected thresholds (5%, 10%, 25%). **B) Time-frequency representation of EEG signal.** Subject average of log-ratios between stimulus and baseline of time-frequency transformed virtual channels, optimized for low and high frequencies. Only the left hemispheric channels are shown. The white square indicates the data points that were included in the combined EEG-fMRI analyses. Average reaction time and stimulus onset are indicated by a white line. **C) Average BOLD response for selected conditions.** Surface projection of average t-map of the first level contrast for left stimulus orientation compared to baseline (left) and the contrast between preferred and non-preferred stimulus orientation (right). **D) Differential voxel response.** The width of the base-line represents all voxel in a specific area (here V1). Those are separated into P_c and nP_c , by contrasting both orientations and splitting the t-value map at 0. From there, a sub-selection for P_c ($t > 0$) can be made by selecting those voxel that expose a positive response to the preferred stimulus compared to baseline and a negative response for the respective other orientation. Again, from positive voxel of $P_c P^+$ those that respond stronger to one feature or the other. Width of each stripe represents the current sub-selection compared to the base-line. Note, that there is a substantial selection of voxel that respond positive to the preferred, but negative to the not preferred stimulus (middle of the plot).

respective first level contrasts expose a clear spatially separable pattern for **A**, **P_c** and **nP_c** , but not **$P_c - nP_c$** . The first three are expected to concentrate spots of highest activation at the occipital pole due to the central fixation of the participants [32, 359], whereas the latter is expected to not expose such a pattern due to the way the contrast was constructed. Both expected spatial patterns have been observed. The left Figure 15 (C) shows the average t-map of the first level contrast for the left oriented stimulus vs baseline. A similar pattern can be observed for all of the aforementioned contrasts as well. As clearly visible, the activation peaks at the occipital pole. On the right hand side of Figure 15 (C) the average t-map for **$P_c - nP_c$** is shown. No clear pattern of any kind of spatial distribution can be observed. Furthermore, Figure 15 (D) depicts respective average voxel subselections per condition. The

lowest level reflects the number of voxel on average over all participants for all voxel within V1. In a first separation, \mathbf{Pc} and \mathbf{nPc} are separated, by contrasting both stimulus features and extracting positive or negative t-values. \mathbf{Pc} (and in reverse \mathbf{nPc}) can be grouped further into voxel that expose a positive response to the preferred stimulus orientation compared to baseline (\mathbf{P}^+) or a negative response (\mathbf{P}^-) and from there on, \mathbf{P}^+ can be grouped into those voxel that have a higher or lower response when presented with the respective *other* stimulus orientation (\mathbf{PnP}^+ or \mathbf{PnP}^-). In addition to $\mathbf{PnP}^+ / \mathbf{PnP}^-$ or $\mathbf{PcP}^+ / \mathbf{PcP}^-$ or $\mathbf{PcnP}^+ / \mathbf{PcnP}^-$; $\mathbf{PcA}^+ / \mathbf{PcA}^-$ can be computed as well, to sub-select those feature specific voxel that are (de-) activated on a general level, since \mathbf{A} contains both stimulus orientations.

Combined EEG-fMRI analyses The feature specific relationship between the EEG and fMRI data has been investigated on a trial by trial basis by means of a general linear model (GLM). Thereby stimulus orientation specific EEG regressors have been related to BOLD signal changes for congruent (preferred: \mathbf{Pc} or \mathbf{P}) and non-congruent (not preferred: \mathbf{nPc} or \mathbf{nP}) combinations of orientation specific data. All analyses have been conducted for primary regions V1, V2 and V3. A separate general linear model (GLM) model was computed for each time-frequency bin derived EEG regressor that was constructed by convolving the EEG response of time-frequency transformed virtual channels with the standard hemodynamic response function as built into SPM12. Task and nuisance fMRI regressors served as control parameters and were fixed. Each β weight for every voxel that was taken into consideration was multiplied by the layer weights for a respective voxel. This procedure was applied separately to low and high frequencies. Furthermore, the procedure was repeated for activation thresholds used to select a respective subset of voxel, such that the most activated 5%, 10% and 25% voxel have been considered. For the main analyses, data was averaged for a time window between 0.1 s and 0.8 s after stimulus onset. Feature specificity was investigated, by contrasting the relationship between preferred orientation responses and non-preferred orientation responses ($\mathbf{Pc} - \mathbf{nPc}$). In order to replicate findings by Scheeringa et al. (2016) [238], the time window was set to 0.1 s to 1.4 s and the \mathbf{Pc} contrast was chosen. Statistical inference was conducted by means of a cluster permutation test [352] for a selected frequency range between 8 Hz and 14 Hz for low and between 50 Hz and 70 Hz for high frequencies, based on the average EEG response (see Figure 15 (B)). Layer specificity was determined by averaging the widest possible frequency window within a significant cluster of the *depth* \times *frequency* result matrix. The depth profile was then tested using an autoregressive rank order similarity (aros) test [353]. It was tested, whether the rank order of the layer profile after averaging over the frequency domain, can be explained better by the data than it could be if the data could not be meaningfully sorted into the respective groups (layers) and would thus be random.

General (de-) activation

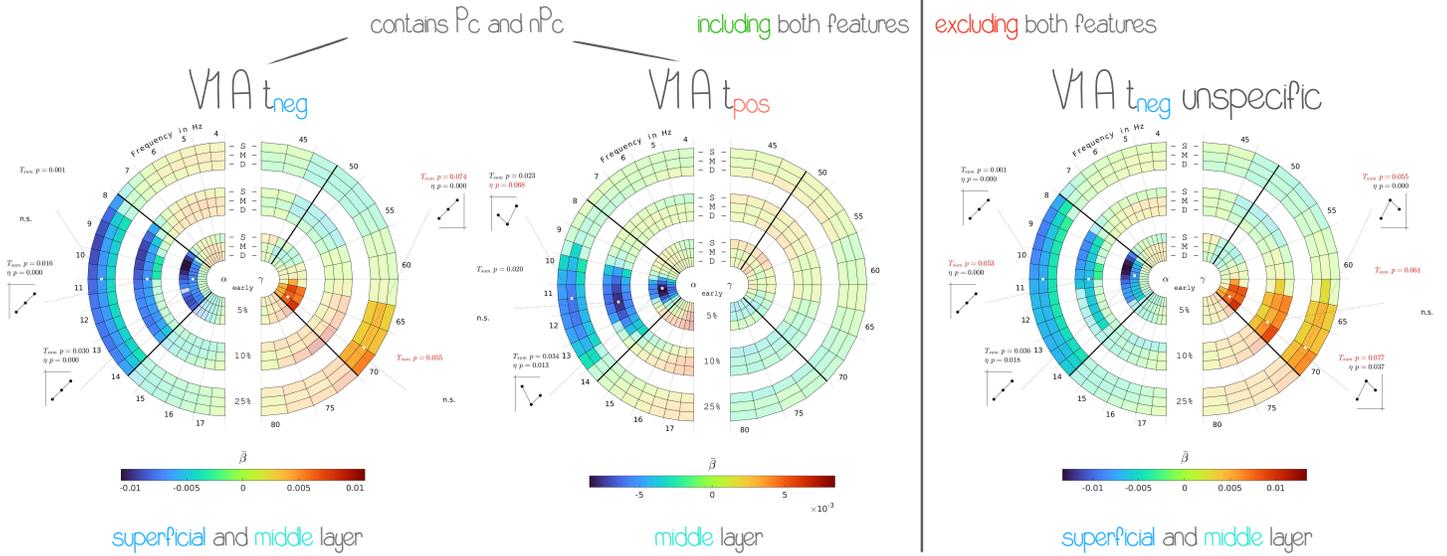


Figure 16: **General (de-) activation.** Result of cluster and aros test for the correlation between α and γ band oscillations for either activated ($A_{t_{pos}}$ or A^+), deactivated ($A_{t_{neg}}$ or A^-) - figures to the left of the line - or deactivated voxel where the average feature specific signal has been regressed out ($A_{t_{neg}}$ *unspecific*) - figure to the right of the line. α band oscillations are significantly negative correlated for both, positive and negative relative changes of the BOLD signal. Respective effects are strongest in superficial and middle layers. Since, A^+ inevitably includes feature specific signal contributions, the correspondence between A^- and A^- *unspecific* indicates a general process related to feature unspecific deactivation in superficial (and middle) layers. Note, that for A^+ only *middle* layer is indicated (two conflicting aros test results only share the middle layer with the commonly highest negative correlation). While A^- seems to reflect a more general deactivation (similar in A^- and A^- *unspecific*), A^+ differs in terms of frequency band width and layer activation profile. This might be explained by a significant contribution of processes related to the separation of preferred and not preferred stimulus orientation. See Figure 19 where A^\pm is combine with feature specific selectivity. A trend level upper γ band response has been found for A^- and A^- *unspecific* with strongest effects in superficial and middle or superficial and deep layers. See also Figure 19 (A).

The spatial distribution of the BOLD response with respect to stimulus orientation and **general activation** compared to baseline (see Figure 15 (C) left), has been interpreted as retinotopic reflection of attention, since the resulting pattern would be typical [354]. Note, that attention has not specifically been manipulated and is used here loosely synonymous with general activation for both stimuli. In order to investigate the general activation or attention effect with respect to the combined EEG-fMRI analysis, all voxel in V1 have been separated into positive or negative compared to baseline for *any* stimulus (A). For both, positive and negative general BOLD activation ($t > 0$ or $t < 0$ respectively), a significant cluster for negative correlations between α band power changes and the top 5% ($p = 0.034$ and $p = 0.03$), 10% ($p = 0.02$ and $p = 0.016$) and 25% ($p = 0.023$ and $p = 0.001$) most (de-)

activated voxel has been found. A significant aros test [353] result for \mathbf{A}^- could be obtained for the top 5% and 10% most deactivated voxel, where superficial and middle layers exposed highest negative correlations with the BOLD signal (both $p < 0.001$). In turn for \mathbf{A}^+ , the aros test reveal a significant layer profile only for the 5% most deactivated voxel ($p = 0.013$) and a trend for the 25% most deactivated voxel ($p = 0.068$) where middle and deep or middle and superficial expose the strongest negative correlations respectively. Additionally, the top 25% average signal of \mathbf{Pc} and \mathbf{nPc} has been regressed out from \mathbf{A}^- in order to reveal "true" feature **un**-specific general deactivation. Two significant clusters for the negative correlation between α band power changes and the BOLD signal have been observed for the top 5% and 25% most deactivated voxel ($p = 0.036$ and $p = 0.001$ respectively) as well as a almost significant cluster for the 10% sub-selection ($p = 0.053$). For those three clusters, a aros test revealed that strongest effect is observed in superficial and middle layers ($p < 0.001$ for the top 10% and 25% and $p = 0.018$ for the top 5% most deactivated voxel). Clusters for negative t values span the entire α spectrum (8 Hz to 14 Hz), whereas cluster for positive t values span only 10 Hz to 13 Hz . In addition to that, a positive correlation with the upper half of the γ frequency spectrum of interest could be observed for \mathbf{A}^- (top 5% $p = 0.074$ and top 25% $p = 0.055$) as well as for $\mathbf{A}^-_{unspecific}$ (top 5% $p = 0.077$, top 10% $p = 0.061$ and top 25% $p = 0.055$).

Contrasting \mathbf{Pc} and \mathbf{nPc} , revealed significant clusters in V1 only for the 25% thresholds in γ . The γ band cluster spans a frequency range between 45 Hz and 65 Hz (25% $p = .019$, trend for 10% $p = 0.067$) and exposed a significant layer profile, such that *deep > superficial > middle* layer ($p < 0.001$ for both). No significant result could be obtained from \mathbf{Pc} or \mathbf{nPc} alone (but see Figure 18). Within the α range clusters span a frequency range of 11 Hz to 14 Hz (5% $p = .033$, 10% $p = .034$, 25% $p = .035$). A significant layer specificity could be observed, such that the relationship between EEG regressors and BOLD signal can be sorted as *superficial < deep < middle* layer (10% $p = .007$, 25% $p = 0.035$) or *deep < superficial < middle* (25% $p < .001$). Feature and frequency specificity have been found in the α band. Thereby, low frequency α exposed a relative decrease across the α range for frequencies between 8.5 Hz and 10 Hz for the correlation with not preferred as compared to preferred voxel in deep and middle layers (deep $p = 0.043$, middle $p = 0.012$). This relationship is reversed for frequencies between 11.5 Hz and 13 Hz and the correlation with preferred as compared to non preferred voxel (deep $p = 0.026$, middle $p = 0.026$). No such frequency specific effect has been found for superficial layers (but see Figure 19). See Figure 17 for a depiction of the results.

Feature and frequency specific (de-) activation has been found to be related with the α but not γ frequency band in both \mathbf{Pc} and \mathbf{nPc} . Signals from either \mathbf{Pc} or \mathbf{nPc} have been analyzed for the fraction of voxel that exposed a relative (de-) activation compared to baseline for \mathbf{Pc} and \mathbf{nPc} (corresponds to \mathbf{A}). Irrespective of the sign of activation of \mathbf{A} ,

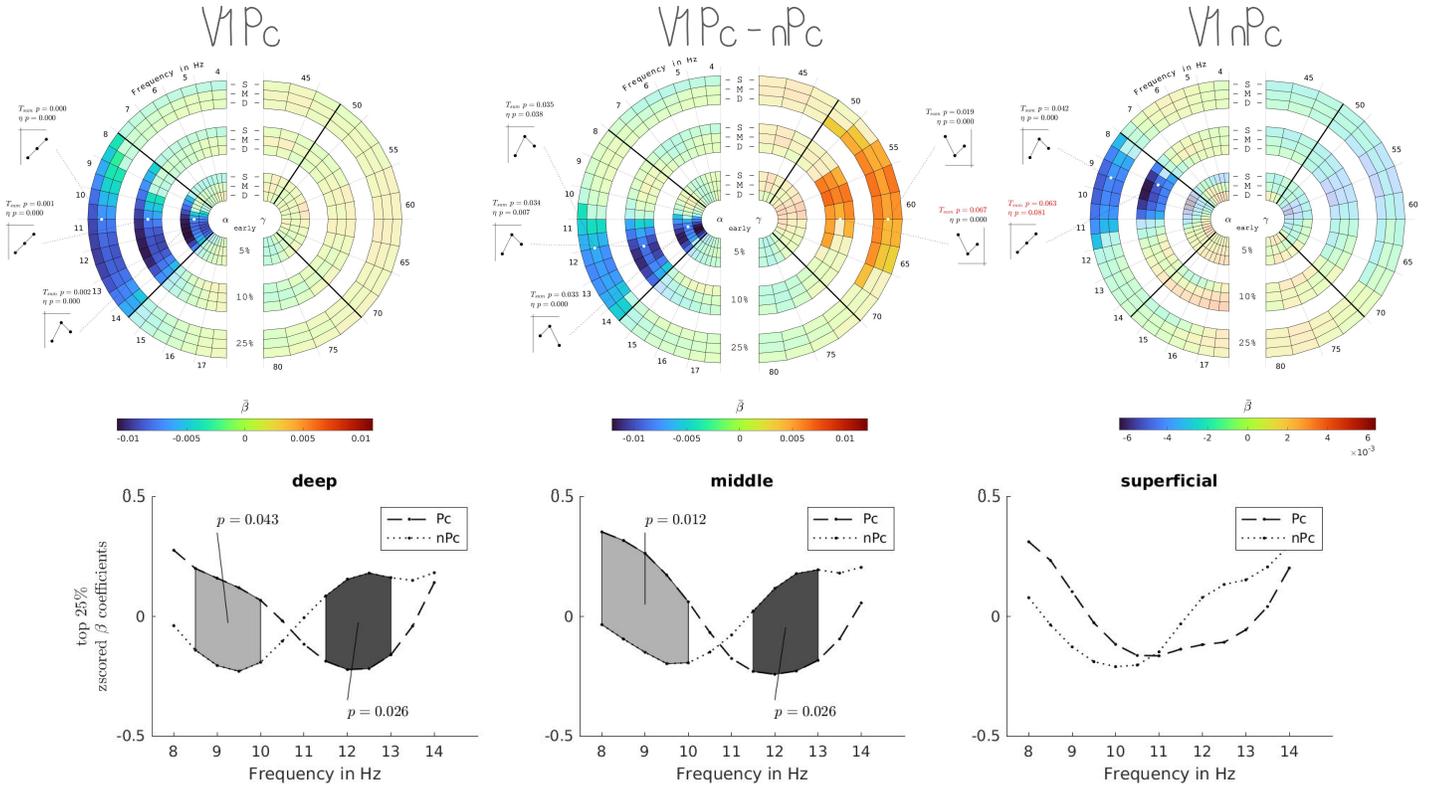


Figure 17: **Feature contrast.** Contrast between the preferred (**Pc**) and non-preferred (**nPc**) orientation for the relationship between EEG regressors and BOLD signal for low and high frequencies in V1 for top most 5%, 10% and 25% active voxel. Top: Highlighted areas indicate a significant relationship between EEG regressors and the BOLD signal, tested using a cluster permutation test [352]. Thick black lines indicate the frequencies of interest that were included for the test. For each cluster an auto-regressive rank order similarity (aros) test [353] has been conducted. Within each cluster, the widest possible range of frequencies has been selected over all three layers and averaged to obtain the layer profile. Significant test results are indicated by the small coordinate system in indicating the profile itself together with the respective p-value as a result of the aros test. The higher the dot relative to the coordinate system, the higher the respective rank order value. Superficial layers are represented closest to the vertical axis. Color mapped regression coefficients reflect the average contribution of one voxel. Bottom: The normalized frequency profile for **Pc** and **nPc** has been contrasted using a two sided cluster permutation test [352]. Significant clusters and p-values are indicated for each layer. Note, that the contrast is shown for the 25% threshold only.

a significant negative correlation between α power and the BOLD signal for **Pc** has been observed for all thresholds (5% $p = 0.002$, 10% $p = 0.001$, 25% $p < 0.001$). For the 10% and 25% percent thresholds, correlations over layer can be sorted, such that *superficial* < *middle* < *deep* layers (10% $p < 0.001$, 25% $p < 0.001$) and for 5% *superficial* < *deep* < *middle* layer ($p < 0.001$). Significant frequency clusters - irrespective of the threshold - span the entire α frequency band (8 Hz to 14 Hz) but are clearly biased towards higher

α frequencies ($> 11 \text{ Hz}$). For **nPc** only a negative cluster for the 25% threshold has been observed ($p = 0.042$), where respective layer signals could be sorted, such that *superficial* $<$ *deep* $<$ *middle* ($p < 0.001$). Here, the cluster spans frequencies from 8 Hz to 11 Hz . A differential comparison between the normalized α frequency profiles revealed differential α correlations in deep and middle layers. High α frequencies (11.5 Hz to 13 Hz) are relatively more negative correlated for **Pc** compared to **nPc** (deep $p = 0.026$, middle $p = 0.026$), whereas lower α frequencies (8 Hz to 10 Hz) are relatively more negatively correlated with **nPc** (deep $p = 0.043$, middle $p = 0.012$). Similar cluster results could be obtained from the correlation between **Pc** or **nPc** voxel activity with α for those that are strictly limited to positive global activation (**A**⁺) or deactivation (**A**⁻). For a detailed overview, including p-values for the lastly mentioned analyses can be found in Figure 19.

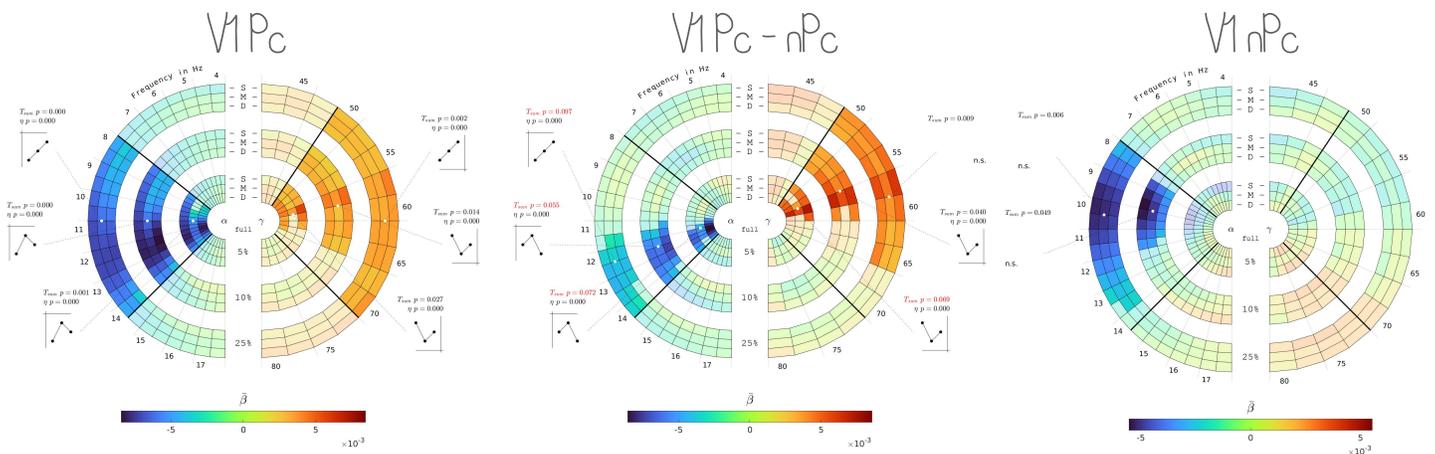


Figure 18: **Full trial analysis** This analysis has been performed as a sanity check in order to replicate Scheeringa et al., (2016) [238]. Relationship between EEG regressors and BOLD signal for low and high frequencies in V1 for top most 5%, 10% and 25% active voxel. Highlighted areas indicate a significant relationship between EEG regressors and the BOLD signal, tested using a cluster permutation test [352]. Thick black lines indicate the frequencies of interest that were included in the test. For each cluster an auto-regressive rank order similarity (aros) test [353] has been conducted. Within each cluster, the widest possible range of frequencies has been selected and averaged to obtain the layer profile. Significant test results are indicated by the small coordinate system in turn indicating the profile itself together with the respective p-value as a result of the aros test. The higher the dot relative to the coordinate system, the higher the respective rank order value. Superficial layers are represented closest to the vertical axis. Color coded regression coefficients reflect the average contribution of one voxel.

The **replication analysis of Scheeringa et al. (2016) [238]** revealed a significant cluster in the α range for 5% ($p = .001$), 10% ($p < .001$) and 25% ($p < .001$) thresholds and in γ range for 5% ($p = .027$), 10% ($p = .014$) and 25% ($p = .002$) thresholds in V1. In V2 significant clusters could be observed in the α range for 5% ($p = .004$), 10% ($p < .001$)

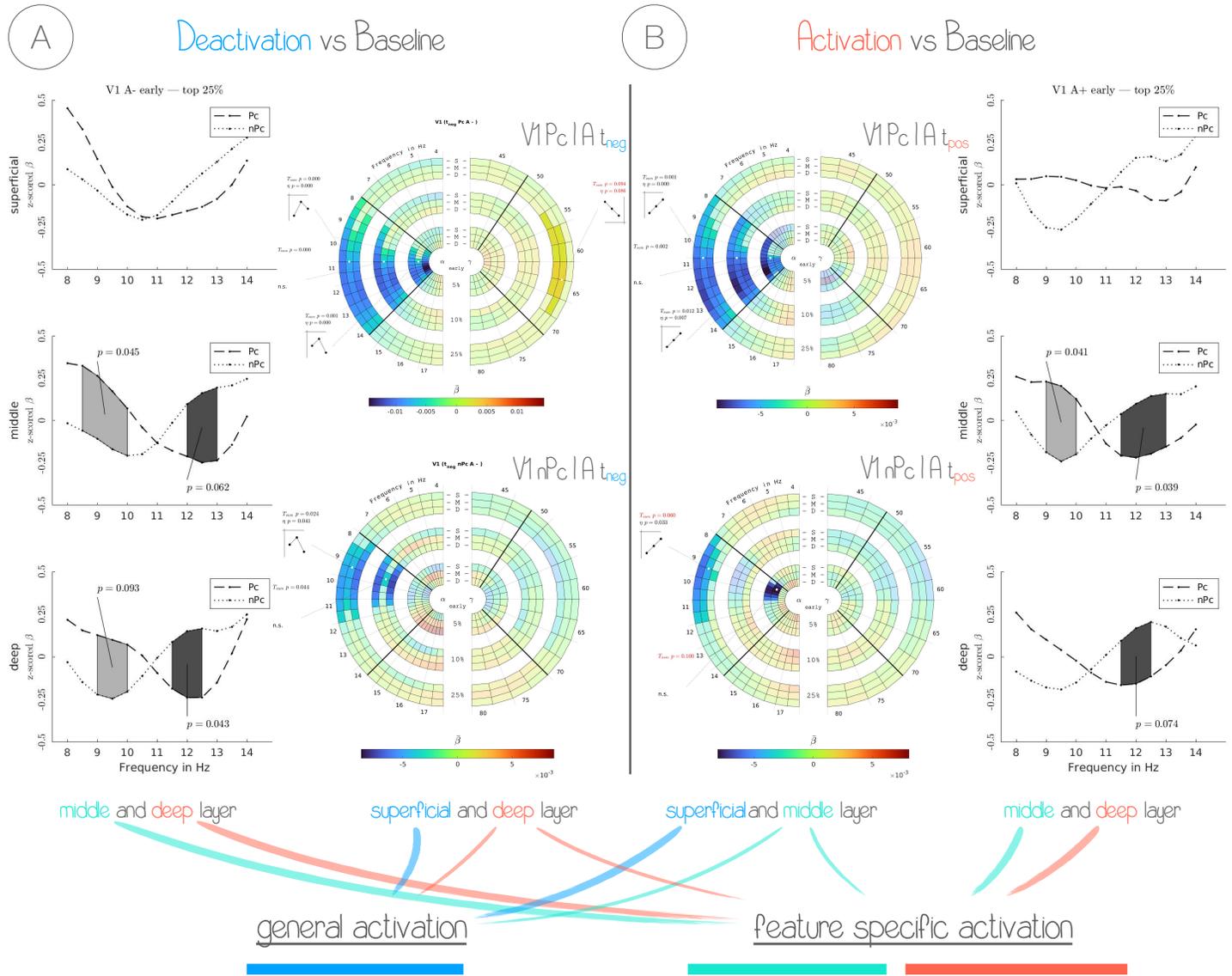


Figure 19: **Feature and frequency selective α (de-) activation.** The negative correlation between α and the BOLD signal has been normalized across the frequency spectrum and was compared between **Pc** and **nPc**. Additionally, sets of feature selective voxel have been intersected with the set of voxel that are globally (in-) active (**A⁺** and **A⁻**). Highlighted colors indicate significant clusters. For each cluster with $p < 0.1$ an aros test [353] has been performed to identify the respective layer activation profile (p-values and shapes are indicated). Global (feature un-specific) α power changes have been found predominantly in superficial layers, whereas feature specific α band decreases have been linked to an increase in the BOLD signal mainly in deep layers. Middle layers could be related to both processes and are hence considered to carry a mixture of both signals. Interestingly, in superficial layers **nPc** seems to express a similar response pattern in **A⁺** and **A⁻**, but **Pc** seems to expose a different response pattern across **A⁺** and **A⁻** indicating a combined attention and preferred feature process. Furthermore, this assumption would explain the differential findings for **A⁺** and **A⁻** depicted in Figure 16 left. In addition to that a trend level γ band response has been observed for **PcA⁻** mostly in deep (feature specific) layers.

and 25% ($p = .001$) thresholds and in γ range for the 5% ($p = .027$) threshold. Significant clusters in V3 could be observed in the α range for 5% ($p = .003$), 10% ($p = .001$) and 25% ($p < .001$) thresholds but not significant clusters for γ . See Figure 18 for a detailed overview.

Discussion

In the present study, laminar level fMRI was combined with simultaneously recorded EEG in human participants. We tested whether feature specific BOLD activity changes - across cortical layers - are related to feature specific frequency power variations in the EEG data as predicted by multiple frameworks [82, 91, 240, 246]. After preferred stimulus orientations have been assigned to each voxel, a general linear model (GLM) was used to assess the relationship between EEG power modulation regressors and the BOLD signal.

Feature processing specific γ We found feature and cortical depth specific effect for trial-by-trial power changes of oscillations in the γ band in V1 for the difference between preferred and not preferred voxel (**Pc** and **nPc**, see Figure 17). EEG power in the γ band (57.5 Hz to 65 Hz) was related positively to the BOLD signal of voxel exposed to their preferred orientation and negative for voxel exposed to the non-preferred orientation, predominantly in deep and superficial layers. In addition to that, findings by Scheeringa et al. (2016) [238], who proved feasibility for relating single trial EEG power to the fMRI BOLD signal, could be replicated by analyzing the full trial period. A strong positive relationship between γ power fluctuations and the BOLD signal in deep and superficial (also middle) layers was found. This has been predicted by animal research [52, 272] anatomical findings [80, 238] and theoretical frameworks [91, 240, 246]. In the present experiment, only the 25% threshold exposed a significant γ correlation, but a trend for the 10% threshold has been observed as well ($p = 0.069$). A possible explanation is the low signal-to-noise ratio (SNR) or low contrast-to-noise ratio (CNR) for this type of experiments. Alternatively, γ band oscillations do not reflect the processing of *specific* features but rather the general process of feature processing which is implicit to the communication through coherence (CTC) hypothesis [240, 246] and communication via nested oscillations (CNO) models [91, 204]. Indeed, only the contrast between both conditions revealed a significant cluster. An analysis of both conditions separately did not. However, a significant cluster for **Pc** for the γ band was found when the full trial was included. What sounds counter-intuitively here is, that if γ is related to specific feature processing - a process completed early in the trial, indicated by reaction times - then why does the correlation between γ and the feature specific BOLD only exposes a significant result when including the "post-decision" period? According to Bonnefond et al. (2017) [91], γ band oscillations are nested into α oscillations that set up a communication channel between task relevant neuronal populations. Since the task was actually to determine a small "wiggly" pattern, the main feature of attention might not have been ori-

entation but instead e.g. "waviness". This means that the actual relevant decision about the stimulus, does not include information about the orientation. Orientation specific processes however would nevertheless be expected on the level of V1, provided those kind of stimuli [30]. Possibly, the SNR requires the full trial to be included in order to make this effect visible within a single condition. On the other hand, the EEG signal exposes a much weaker and less clear signal in the γ band for the late period (see Figure 15). Conversely, full trial length results for γ appear much more broadband. Hence, it can be assumed (based on EEG γ power and BOLD correlation) that the full trials includes additional cortical activity that exceeds feature specific responses. Post-stimulus but pre-response γ power changes in one condition alone do not sufficiently explain respective changes in the BOLD signal. Hence, a plausible explanation would be that γ band power does not reflect specific *feature* activation but feature *determination* (i.e. separation or processing). More surprising here is that γ band oscillations are correlated (on trend level) with general BOLD deactivation, especially for preferred stimuli (see Figure 16 and Figure 19) While one interpretation would be that this correlation just reflects the general relationship between γ and BOLD [238, 360, 361], the fact that this is not observed for generally activated voxel makes this assumption debatable. However, γ band correlated with negative BOLD responses for directed attention could be explained by the "true" underlying computations that become visible after stripping away interaction effects by general activation (attention). If γ band oscillations are nested in coherent α (as described above) in order to allow for signal transmission which is gated by an attentional α component [91], then attended preferred feature signals are composed of a signature of transmission α , attention α and feature γ . While attention could be demonstrated to enhance early visual processing [362] - probably mediated by low frequency oscillations [363] - low frequency *suppression* effects seem to be much more dependent on the actual task and stimulus material [233]. Signals of generally deactivated voxel and voxel coding for the not preferred stimulus features would hence lack the "boost" from attention α and their effects on γ . The missing γ band correlation in \mathbf{PcA}^+ could thus be a consequence of attention processes overshadowing the feature process which however become visible when looking at the generally deactivated voxel, because supposedly the respective attention suppression has less effect than the attention facilitation of generally activated voxel.

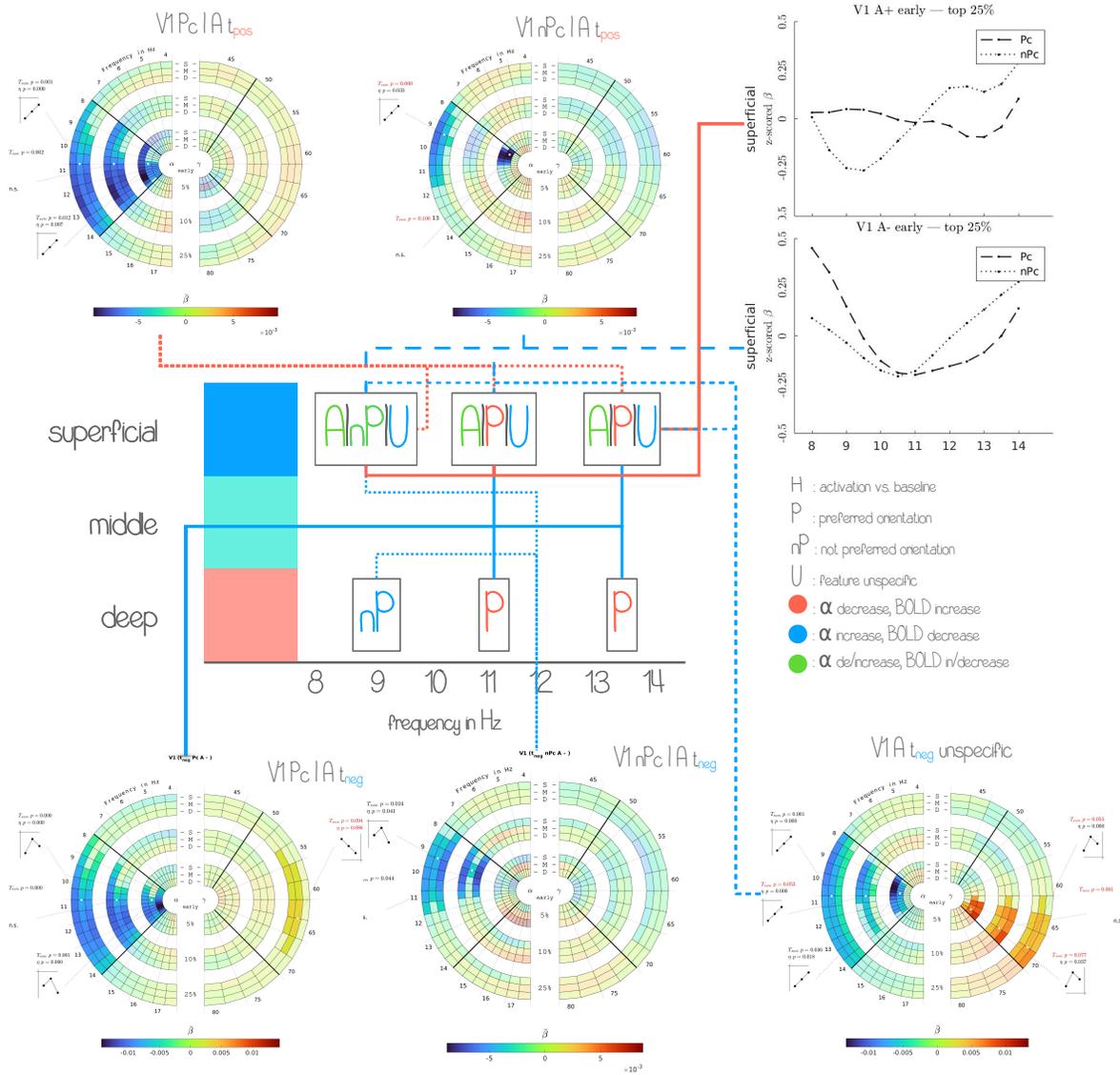


Figure 20: **Feature, layer and BOLD (de-) activation specific α oscillations** The Figure summarizes selected findings (see "Results" on page 80). Thereby, attention related α power changes are predominantly linked to a general BOLD signal modulation mainly in superficial layers, whereas more feature specific processing related α -BOLD negative correlations are located in deep and superficial layers, with a differential profile for preferred and non preferred stimulus orientations in deep and middle layers (high frequency α decrease, low frequency α increase and vice versa for not preferred stimuli). In general the feature specific α related processes are mainly attributed to deep layer BOLD activity (positive *and* negative). Anatomical and functional attributions are depicted in the center. Even though frequency selectivity could only be found in deep and middle layers, the modulation in superficial layers seems to be dependent on A^\pm . Especially the modulation of the frequency profile of Pc seems to drive this difference which needs to be verified in future analyses.

Generally speaking, positive correlations between EEG γ power and the fMRI BOLD signal are a well established finding [238, 360, 361]. Simultaneous recordings of local field potentials (LFP) signals and fMRI in animals suggest a close relationship between neural activity reflected by LFPs in the γ band and the BOLD signal [237, 364]. The EEG signal is closely related to LFPs as well [365–367]. Hence, the relationship between EEG and simultaneously recorded BOLD signal, can serve as a proxy to investigate the relationship between cortical oscillatory patterns and underlying neural activity in humans. The here presented novel approach adds laminar level resolution to simultaneous EEG-fMRI which gives rise to the opportunity of relating oscillatory brain activity to the cortical architecture and underlying neuronal activity and test predictions derived from current cortical computation frameworks. Cortical γ activity has been linked to routing cortical information [240], stimulus features themselves [368] and spike time coordination [369]. Here we find a feature specific γ band increase in the correlation with the BOLD signal, for voxel exposed to their preferred stimulus orientation over not preferred in V1. This effect could be observed for the 25% and on trend level for the 10% threshold. Later layer activation profile analyses revealed highest correlations with the laminar BOLD in deep and superficial layers. Those findings have been predicted by previous literature [52, 91, 238], but deep and superficial layers have been expected to expose the reverse correlation pattern (*superficial* > *deep*). Nevertheless, anatomical γ related feed-forward pathways have been attributed to deep and superficial layers [80, 82, 246]. Further analyses must investigate the exact connectivity between (and within across layers [265]) in order to understand exact computational pathways.

For α band oscillations we observed a much more complex response pattern for different conditions. Thereby, negative effects between α band power changes and the BOLD signal could be separated into attention (or general activation) and feature (and frequency) specific processes. The close relationship between electrophysiological and BOLD signals has been demonstrated previously. Negative correlations with BOLD response could be shown for θ [370, 371], α [238, 372, 373] and β [303, 374, 375] oscillations.

Attention specific α The analysis of the BOLD signal distribution for each stimulus feature and both combined compared to baseline, revealed relatively clear separation between positive and negative t values within the respective contrast (see Figure 15 (C) left). Since this pattern can be observed likely as a consequence of retinotopic (de-) activation in primary visual regions driven by spatial attention [354], generally activated or deactivated voxel might to some extent reflect directed attention [56]. Since it is hypothesized, that separate processes for attention and feature related activity in the α band should be possible to identify [91], those processes must constructively overlap for voxel at attended sites that are presented with their preferred feature since this would reflect the actual ongoing process. In addition to that and increase in α power over not attended receptive field locations is hypothesized

to perform some sort of suppression [224] or at least "neglecting" [176] operation. Indeed we found that decreased α power is related to increased BOLD activity in activated and deactivated voxel (note: more α power in deactivated voxel is related to less BOLD activity). However, the frequency bandwidth between correlations with deactivated and activated voxel is different. While deactivated voxel expose a negative correlation with broadband low frequency activity, the negative correlation with activated voxel is limited to a range between 10 Hz and 13 Hz . Since positive activated voxel might contain strong signal contributions from feature specific voxel that are attended, respective feature specific activity has been removed from the negative BOLD signal. The resulting negative correlation pattern with α is very similar to the negative correlation between α and general deactivation. In addition to that, respective negative BOLD correlations have been found strongest in superficial (and middle) layers, whereas the correlation with activated voxel has been located primarily to middle and deep layer activity. This difference is likely not driven by a SNR related issue, since the layer profile between positive and negative responses and correlations is highly different. A SNR bias due to superficial draining veins that have an effect on the BOLD signal [273], would likely affect positive *and* negative responses similarly (at least in terms of layer bias). A more likely explanation is that a general deactivation is reflected predominantly in superficial layers by the α band. Findings related to BOLD deactivation support this view. Since positive BOLD signal can be interpreted as a proxy for ongoing neuronal processing [32] and attention has been demonstrated to be on one hand related to targeted α band changes [251] and negative relative BOLD signal changes [355] a potential inhibiting relationship between α and BOLD could very well be expected as well. It has been reported that enhanced negative fMRI default mode activity was correlated to higher spontaneous EEG α activity [376]. Here we find a negative correlation between α power and negative BOLD response as well. Moreover, an eyes open eyes closed experiment using ECoG in humans, revealed that during the eyes open condition respective negative α power deflections were mostly related to superficial layers [270]. This more or less target-less but directed attention [56] process would be somewhat similar to our experiment, where directed attention was required to perform the task. In addition to that, Haegens et al. (2015) [271] related superficial α band activity to a more general process of sensory parsing in primary sensory regions. The discrepancy between general activation related to superficial and findings on "true" attention modulation from animal models [52] mostly related to deep layer activity, possibly can be resolved by the respective measure, as both, Haegens et al. (2015) [271] and Halgren et al. (2019) [270], argue that LFPs report spurious deep layer activity related to attention, due to volume conduction.

Feature specific α The inhibitory nature of α oscillations to suppress (or neglect [176]) unwanted information is a well established hypothesis [173, 224, 231, 377, 378]. Here we could show that an decreased α power is linked to increased BOLD activity in voxel that process their preferred orientation, whereas an increase in α power has been found to be linked to lower BOLD activity in voxel that were exposed to non-preferred stimuli in superficial and deep layers. The contrast between **Pc** and **nPc** controls for general activation "background" effects (e.g. receptive field specific or spatial attention [354]) because e.g. spatial attention is not hypothesized to affect stimulus feature processing per se. However, some theoretical frameworks predict multiple sources of α that either set up neuronal communication or gate cortico-cortical information flow [91, 148]. Applying such a framework would predict that attention specific α and feature processing specific α contribute jointly to the process. Hence, the contrast between **Pc** and **nPc** controls only for e.g. spatial attention but not interaction effects between spatial attention and feature related processes. Indeed, the negative correlation pattern between α for the contrast between features (**Pc** – **nPc**) exposes a similar frequency band limitation as observed for **A⁺** (see Figure 16 center and Figure 17). The layer profile for **Pc** – **nPc** locates the strongest negative effects to superficial and deep layers. A more specific analysis of α for **Pc** and **nPc** separately for either **A⁺** or **A⁻** (see Figure 19) revealed however a feature specific activity pattern in the α band that has been linked predominantly to deep and superficial layers. A separate analysis of **Pc** and **nPc** (see also Figure 17) indicates that negative correlations for oscillations in the α band exposed a peak lower frequency within the spectrum (8 Hz to 10 Hz) for not preferred as compared to preferred orientations (10 Hz to 14 Hz). A permutation test revealed *where* the respective α frequency profiles for **Pc** and **nPc** differ. Predominantly in deep and middle layers, the frequency profile exposed a higher negative correlation between α and the BOLD signal for **Pc** in the upper α band (11 Hz to 13 Hz) as compared to **nPc**. This pattern was reversed for the lower α band (8 Hz to 10 Hz) (see Figure 17). Most notably, this exact pattern was observed for generally activated *and* deactivated voxel. Even though, a difference with respect to frequency-specificity in superficial layers could not be observed for activated or deactivated voxel, the respective frequency profiles expose a striking dis-similarity. While the overall correlation for **Pc** and **nPc** is very similar in superficial layers in deactivated voxel, this is not necessarily the case for the generally activated selection. Interestingly, the response for **nPc** shares high similarities for activated and deactivated voxel in terms of α frequency profile, however the patterns for **Pc** seemingly differs (see Figure 20 right). As described above, attention related processes could be mainly linked to superficial α activity. The contrast between frequency profiles of **Pc** and **nPc** for activated and deactivated voxel in turn suggests a feature processing related role of deep layer α where processes with respect to preferred and not preferred orientation expose frequency specificity (see also Figure 20). This would support the idea that α might play a key role in attended feature selection [224], but furthermore can be related to feature specific processes and network communication [255], which is in line with theoretical frameworks taking all those aspects into account

[91, 291]. Feature selective α band activity has been predicted [91]. However, a differential feature *and* frequency selective α oscillation, establishes a very new and interesting finding. Potential follow up experiments specifically need to target this frequency dissociation with respect to relative power changes and connectivity. Furthermore, it needs to be investigated, whether indeed superficial layer α in \mathbf{PcA}^+ and \mathbf{nPcA}^\pm is statistically different. One approach would be to compute the similarity between frequency profiles in superficial layers between \mathbf{A}^+ and \mathbf{A}^- for \mathbf{Pc} and \mathbf{nPc} separately. According to current theoretical frameworks, neuronal commutation is established in feedback direction, by synchronized α band activity between regions of interest, controlling the flow of information. Targeting specifically neuronal populations exposed to their preferred stimulus orientation, would thereby result in a neuronal population (feature) specific α decrease reflecting feature specific dis-inhibition, similar to what we could show.

Relationship to theoretical frameworks While γ oscillations are on one hand considered to be a direct consequence of stimulus features and the neuronal architecture [189, 368] other theories point towards a role of γ in functional network communication [240]. If however γ oscillations are phase locked to the α cycle in local computations [91, 259, 379], increased γ power might reflect the process of information processing itself as it occurs. Indeed the correlation between γ power and feature specific BOLD signals is only present for \mathbf{Pc} during the full trial and crucially not for \mathbf{nPc} . Feature specific processing (eliciting high γ power) might indeed be set up by low frequency (i.e. α) oscillations, dis-inhibiting voxel preferably processing a certain stimulus orientation. Phase-amplitude coupling between feature specific α and γ oscillations needs to be investigated in future research. However, the absence of γ oscillations in \mathbf{nPc} at least indicates the presence of some form of inhibitory feedback process. This idea would further explain, why the deep layers for $\mathbf{Pc} - \mathbf{nPc}$ expose a higher correlation with the feature specific BOLD signal, than superficial or middle layers. Since stimulus induced γ is mostly observed in superficial layers [52, 91, 238], it would be expected that this pattern could be observed in \mathbf{nPc} voxel as well. This process would thereby flow into feed-forward direction [52, 91]. Nevertheless, it has been found that top down controlled activity is mainly routed through deep layers [34]. A feedback directed inhibitory signal would lead to a BOLD signal decrease which would be particularly present in deep layers, and would hence produce a stronger (positively) correlated signal in those layers (less γ , less BOLD) for the contrast. After an initial stimulus evaluation period, feedback connections would "lock" the perceived stimulus orientation by dis-inhibiting neuronal populations in \mathbf{Pc} and inhibiting neuronal populations in \mathbf{nPc} [179]. In this model, γ oscillations changes could be interpreted as markers for ongoing stimulus processing in a selected dis-inhibited subsection within primary visual regions, which would be in line with findings to α / γ phase-amplitude coupling [380].

Figure 20 summarizes main findings and indicates respective links to anatomical layers and functional roles. General activation (or attention) related processes, are mainly related to α power changes in superficial and middle layers. This can be observed for the general activation, strictly positive and strictly negative activation (deactivation). Furthermore, stimulus feature related α band correlations have been found to be strongest in deep and superficial layers. Thereby, preferred features are reflected by a high α negative correlation increase and not preferred features by a low α increase in negative correlation. In a nutshell, those findings can be interpreted as two differential α systems for feature (deep layers) and spatial (superficial layers) attention or neuronal recruitment systems.

We reached our main goal of demonstrating the feasibility of relating feature specific EEG power fluctuations to cortical activity in feature specific populations on the level of cortical laminae. This enabled us to investigate predictions derived from animal research using LFPs [52] or recently developed theoretical frameworks [91, 246]. It remains to be investigated whether the directionality of the information flow across cortical regions (γ feed-forward, α feedback) adds further evidence to the respective frameworks. Feasibility of such an analysis has already been demonstrated [265]. This would allow to test the presumed feedback connectivity in α from higher to lower order visual regions [52, 91] predominantly in superficial and deep layers, especially in the contrast condition. As a result, inhibited feed-forward information flow expressed in γ are likely predictable from the correlation with α power and BOLD connectivity. Additionally, an intra-area correlation of α with the BOLD signal for deep and superficial layers would be predicted [265]. In general this study implies that combined EEG-fMRI can be used to study the relationship between oscillatory brain activity and the cortical architecture. However, future research might not only employ different modalities to confirm our findings (e.g. using intra-cranial electrodes), but might as well extend the possible range of experimental paradigms and analysis strategies. Intra and inter regional connectivity could thereby put current theoretical frameworks to the test.

Auto-regressive rank order similarity (aros) test (pre-print)

Clausner, T., & Gentili, S. (2022). Auto-regressive Rank Order Similarity (aros) test. bioRxiv.

<https://www.biorxiv.org/content/10.1101/2022.06.15.496113v1>

Abstract In the present paper we propose a non-parametric statistical test procedure for interval scaled, paired samples data that circumvents the multiple comparison problem (MCP) by relating the data to the rank order of its group averages. Using an auto-regressive procedure, a single test statistic for multiple groups is obtained that allows for qualitative statements about whether multiple group averages are in fact different and how they can be sorted. The presented procedure outperforms classical tests, such as pairwise conducted t-tests and ANOVA, in some circumstances. Furthermore, the test is robust against noise and does not require the data to follow any particular distribution. If A is a data matrix containing N observations for k groups, then the test statistic η can be computed by $\eta = \sum_{i=1}^N f(A_i, \mathbf{s})/N$, where \mathbf{s} is a vector of length k containing the average for each group, transformed into unique rank values. This statistic is compared to the distribution D , obtained by Monte Carlo sampling from the permutation distribution. It will be demonstrated that D can be described by a normal distribution for a variety of input data distributions and choices for f , as long as a set of criteria is met. Comparing η to the permutation distribution controls the false alarm (FA) rate sufficiently, since the exact p -value can be estimated [381]. Multiple examples of possible choices for f will be discussed, as well as detailed descriptions of the underlying test assumptions, possible interpretations and use cases. All mathematical derivations are supported with a set of simulations, written in Python that can be downloaded from <https://gitlab.com/TommyClausner/aros-test> together with an implementation of the test itself.

Introduction In this paper we propose a paired samples, non-parametric statistical procedure on the basis of a permutation test, which aims to circumvent the multiple comparison problem (MCP) by combining multiple group averages into a single statistical value. The MCP arises from the fact that in order to control the FA rate for comparing more than two groups, the critical α level needs to be adjusted. Commonly accepted levels for FA rates are less than 5% or 1%. This means that the null hypothesis (H_0) was falsely rejected in less than 5% or 1% of all cases. As the number of groups k increase, the number of tests n_t increases with $n_t = k!/(2(k-2)!)$. If no adjustments to the critical α was made, the FA rate increases from ≤ 0.05 to $\leq 1 - (1 - \alpha)^{n_t}$, which for three groups means the FA rate becomes ≤ 0.14 . This issue is commonly referred to as multiple comparison problem. One common strategy is to adjust the α level (lower it), until a satisfactory control for the FA

rate is achieved ¹². Adjusting the α level however comes at the cost of potentially reducing statistical power, that is the sensitivity of the test, and thus the rate at which the so called type II errors (falsely accepting H_0) are produced increases.

A widely known test used for statistical analyses of more than two groups is the analysis of variances (ANOVA). The one-way ANOVA, for instance, uses an F -test to relate the variances within and between groups to each other. However, a significant result (i.e. rejecting the H_0 , which states no difference between the group means) only indicates a difference between any subset of the means. Thus, a significant result is often followed by a pairwise comparison, which again leads to the aforementioned MCP.

Furthermore, classical tests, such as z -tests, t -tests or F -tests often require the underlying data to be normal distributed or at least that the parameters of the underlying distributions are known. Instead of relying on knowledge of the underlying distribution, the distribution under H_0 can be estimated by a permutation procedure, which can be applied to all of the above classical tests. This procedure is well established for a variety of use cases [352, 382, 383] and allows for an exact estimation of the p -value [381]. The fundamental idea behind permutation tests is that if there is no difference between multiple averages, then the members of each group can be exchanged, because they stem from the same distribution (H_0). Commonly, members of each group are shuffled between the groups and the respective test statistic of interest is computed. By repeating the procedure a large number of times (ideally by iterating through all possible combinations), the permutation distribution under H_0 is obtained [384]. In a last step the test statistic of the original data is compared to the permutation distribution and the p -value is computed as the fraction of all values of the permutation distribution that exceed the test statistic computed from the original data [381]. Since the p -value is computed by Monte Carlo sampling from the permutation distribution, no knowledge about the underlying parameters is required [385]. Thus, permutation tests belong to the family of non-parametric tests.

The proposed auto-regressive rank order similarity (aros) test is constructed as a permutation test and thus belongs to the family of non-parametric tests as well. Additionally, the MCP is circumvented by condensing the relationship between the data and each group average into a single statistic. Precisely, the group data for each paired observation is treated as vector that is related to the vector of the rank order of the average group value, by means of a similarity metric. Thus, the test statistic η can be seen as the average similarity between the observations and the rank order of the averages. It is tested whether this similarity is significantly greater than the average similarity within the permutation distribution under H_0 .

¹²There is a large variety of strategies, which are not discussed here.

The aros test explained The aros test is a non-parametric statistical hypothesis test for paired sampled, interval scaled data for multiple group average comparison. However, as compared to conventional tests on interval scaled data, the aros test does not rely on the difference between the means of multiple distributions, but rather on the relationship of each paired observation within the data to an ordinal scaled profile or shape derived from either the group averages or an external source. If we assume $k = 3$ groups with averages μ_a, μ_b, μ_c , the standard procedure for classical tests would be to pairwise compare those three groups in order to obtain the qualitative and quantitative relationship between those averages. However, when applying the aros test, this question is reduced to the qualitative relationship of how the group averages rank relative to each other and how well the data explains this relationship. This means that if we observe a ranking of the means such that $\mu_b < \mu_c < \mu_a$, auto-regressive rank order similarity tests, whether this relationship can truly be justified by the data (as compared to determining, whether the pairwise difference between those means is zero in a more traditional setting). In other words it is tested, whether - on average - each observation expresses a similar relationship between each group or condition as expressed by the group averages. The most general formulation for the test statistic η can be written as:

$$\eta = \frac{\sum_{i=1}^N f(A_i, \mathbf{s})}{N}$$

where \mathbf{s} is the vector of group averages, transformed into a set of unique, evenly spaced rank values, such that $\mathbf{s} = [s_1, \dots, s_k]$. For the aforementioned relationship $\mu_b < \mu_c < \mu_a$, one possibility would be to set $\mathbf{s} = [3, 1, 2]$. The function f relates each observation $A_{i \in [1, \dots, N]}$ (the values for each observation in all conditions) of length k and \mathbf{s} separately. η is obtained by averaging the results from $f(A_i, \mathbf{s})$. Thereby, the function f can be freely chosen, as long as f relates A_i and \mathbf{s} in terms of similarity. Examples for f might be the least square solution to $\mathbf{s} = A_i \beta_i$ for β_i , the correlation coefficient, the cosine similarity or the explained variance. In this paper we will focus on the least square solution to $\mathbf{s} = A_i \beta_i$. To solve for β_i we denote:

$$\hat{\beta}_i = (A_i^T A_i)^{-1} A_i^T \mathbf{s}$$

Since A_i and \mathbf{s} are vectors, we can rewrite $A_i^T A_i$ as the dot product of A_i with itself and $A_i^T \mathbf{s}$ as the dot product between A_i and \mathbf{s} . Hence, we denote:

$$\hat{\beta}_i = \frac{A_i \cdot \mathbf{s}}{A_i \cdot A_i}$$

By doing so, the test statistic η for $f(x, y) = (x \cdot y)/(x \cdot x)$ can be computed in the following way:

$$\eta = \frac{\sum_{i=1}^N \frac{A_i \cdot \mathbf{s}}{A_i \cdot A_i}}{N}$$

In this case η can be interpreted as the average fit of the data to the rank order shape. If we assume, for example, $k = 3$ groups with means $[\mu_a, \mu_b, \mu_c]$ and an average rank order shape of $\mathbf{s} = [3, 1, 2]$, η can be interpreted as the average fit of each observation A_i to that shape, or in other words, how well the data explains sorting the means, such that $\mu_b < \mu_c < \mu_a$. Note, that we cannot state how much different each mean is from each other, but how much the data supports this rank order. As previously mentioned, f can be tailored to the specific needs of the respective research question. To the authors opinion however, choosing $f(x, y) = (x \cdot y)/(x \cdot x)$, can be applied to a variety of research questions, and gives rise to a straightforward interpretation. Nevertheless, this specific choice for f is limited by the fact that if one observation A_i is zero for all groups, then f cannot be computed and thus η would be invalid. Hence, f needs to be chosen such that it is defined for each observation and each permutation.

It needs to be pointed out that some of the proposed choices for f are magnitude free (i.e. correlation coefficient, cosine similarity and explained variance), whereas other choices, such as $f(x, y) = (x \cdot y)/x \cdot x$ are not. Magnitude free in this context means that the absolute value of η does not depend on the absolute values of either the data A or the values of \mathbf{s} . However, due to the fact that the permutation procedure is applied in a similar fashion, the bias resulting from the absolute values of A (and \mathbf{s}) affects the estimation of η and the permutation distribution equally and thus does not affect the hypothesis test itself.

So far the general idea behind obtaining the test statistic, as well as multiple possible choices for f have been discussed. The actual hypothesis test has been neglected so far. Generally speaking the aros test is a test on the null hypothesis (H_0). As previously mentioned, this test is performed by Monte Carlo sampling from the permutation distribution under H_0 and comparing the initially obtained η to this distribution. In a first step the number of permutations I is defined. I should be a relatively high number (e.g. $I = 10000$). Since for each permutation step an average over N observations needs to be computed, the maximum number of all possible permutation steps for k groups (and thus $k!$ possible permutations per observation) can be computed as $I = k!^N/N$. Where computationally feasible, the computation of the exact permutation distribution D under H_0 (all possible combinations) should be preferred. Otherwise I samples (ideally, but not necessarily, without replacement [381]) need to be obtained by means of Monte Carlo sampling. During each iteration $i \in [1, \dots, I]$, the data A is permuted across groups for each observation separately. Then, the test statistic is computed and added to D . After performing I iterations, D contains I values for the test statistic under H_0 , to which the initially obtained η is compared. The p -value is obtained by computing the fraction of all values in D that exceed η . Figure 21, illustrates this.

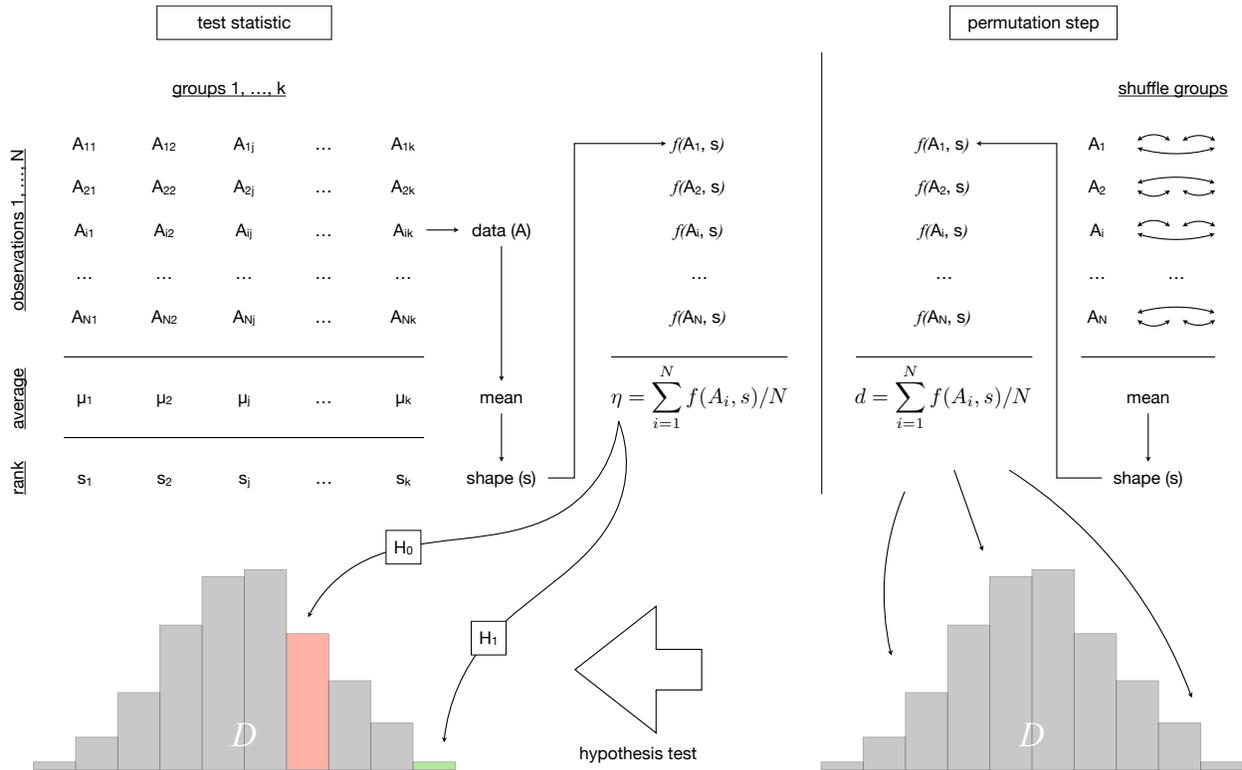


Figure 21: **Procedure of the aros test.** First, the initial test statistic η is obtained by averaging the data over groups and transforming the result into rank values forming the shape s (upper left). Afterwards, the data A is related to s using the function f (upper middle). This procedure is repeated I times to form the permutation distribution D (lower right), except that the data is shuffled over the data columns for each observation separately (upper right). Lastly, η is compared to D (lower left). This is achieved by computing the p -value as the fraction of the values in D that exceed η and compare it to the critical α -value (e.g. 0.05).

During the description of the test procedure, H_0 has been mentioned, but was only implicitly defined. Derived from the general assumption of permutation tests, it is important that the data is exchangeable under H_0 . This means that the joint probability distribution (under H_0) remains the same, irrespective of the order of the single values [352]. Less formally speaking, this means that if there was no difference between multiple groups of values, the values in each group could be interchanged, because they can be assumed to stem from the same distribution. The exchangeability assumption is tightly linked to H_0 itself. If η does not depend on the grouping of the data (H_0), then η would be expected given D , hence:

$$H_0 : \eta = \mathbb{E}[X_D]$$

Under H_0 , \mathbf{s} is derived from A , where each value in $A_i^{[1, \dots, k]}$ could be exchanged. Hence, \mathbf{s} is the result of randomly arranged group labels and its respective average relationship with A - expressed by $\sum_{i=1}^N f(A_i, \mathbf{s})/N$ - could not be predicted by any single A_i or \mathbf{s} . Thus, each value within D could be seen as the average of independent random values. In the "Derivations" section it will be shown that D can be approximated by a normal distribution:

$$\mathcal{N}(\mu, \sigma^2), \text{ with } \mu = \mathbb{E}[X_D] \text{ and } \sigma^2 = \mathbb{E}[(X_D - \mu)^2]$$

As described above, the FA rate is controlled and the hypothesis test performed by estimating the exact p -value \hat{p} and comparing it to the critical α -value [381]:

$$H = \begin{cases} 0, & \text{if } \hat{p} = \frac{1 + \sum_{i=1}^I D_i \geq \eta}{I+1} > \alpha \\ 1, & \text{otherwise} \end{cases}$$

Since the p -value can be computed directly from D as the fraction of values in D exceeding η and comparing it to the critical α -level, the false alarm rate is controlled sufficiently.

Note that D can be approximated by a normal distribution (see "Derivations" for additional details) and thus D could be standardized: $Z = (D - \mu_D)/\sigma_D$ and η expressed in terms of the number of standard deviations it is different from the mean of the permutation distribution (z -scoring).

A special case of aros tests While the standard aros procedure is performed using a shape derived from the data average (data driven), a special case arises if the shape can be derived from an external source. If possible, \mathbf{s} can be obtained by source *independent* of the data (e.g. previous research or derived from the hypothesis). Hence, H_0 changes from asking whether a specific arrangement of the data produces a specific rank order of averages as likely as any other rank order obtained from shuffling the data, to whether under the exchangeability assumption the original arrangement of the data explains a *specific* shape better than any other arrangement of the data. Then, η would be computed based on this *specific* shape. Hence, the shape would not be derived from the average rank order of the data, which biases the estimation of the permutation distribution. By definition, the independence of \mathbf{s} is strictly necessary. In order to avoid "double dipping" \mathbf{s} cannot be derived from the mean and used as *specific* shape of interest. It must be derived from some source other than the data, in order to not violating the independence assumption of \mathbf{s} . However, if \mathbf{s} can be derived from some external source, not violating the independence assumption, then statistical power for this *specific* shape can greatly be increased (see "Comparison of test power"). Furthermore, it is not strictly necessary anymore to require \mathbf{s} to be a unique set of equally spaced values, as the probability for any \mathbf{s} to occur does not need to be uniformly distributed anymore (because there is only one shape). Technically this strips

the aros test of its auto-regressive nature. Nevertheless, the general principle would remain similar enough to view this as a special case of the aros test. Please see "Discussion" for a suggested application, where the standard and special case are combined.

Derivations The derivation is almost immediate given the statement of the central limit theorem.

Theorem 1 (Central limit theorem[386]). *Let X_1, X_2, \dots be i.i.d. with $E[X_i] = \mu$, $var(X_i) = \sigma^2 \in (0, \infty)$. Let $\bar{X}_n = X_1 + \dots + X_n/n$. Then, as $n \rightarrow \infty$,*

$$(\bar{X}_n - \mu) \frac{\sqrt{n}}{\sigma}$$

converges to the standard normal distribution $\mathcal{N}(0, 1)$

From this theorem and the definition of η :

$$\eta = \frac{\sum_{i=1}^N f(A_i, s)}{N}$$

we can directly conclude:

Corollary 1. *Assuming that $f(A_i, s)$ forms a probability distribution with mean μ and variance σ^2 , then, for $N \rightarrow \infty$, the distribution of η converges to a normal distribution $\mathcal{N}(\mu, \sigma^2/N)$*

Given this result one can conclude the distribution of η statistics can be approximated by a normal distribution.

Now that the distribution of η statistics follows a known probability distribution, we can conclude that the probability distribution $f(X_\eta|D)$, that is the distribution of the test statistic given the permutation, is well defined. Finally, we only need to prove that in general the permutation test controls the FA rate [352]. We only need to show that the α level of the permutation distribution is the same as the real α , i.e. $P(\text{reject}H_0|H_0)$.

Theorem 2. *in a permutation test, the probability $\alpha = P(\text{reject}H_0|D, H_0)$ is equal to $P(\text{reject}H_0|H_0)$*

Proof. We rewrite $P(\text{reject}H_0|H_0)$ to be conditioned on D , as

$$\sum_D P(\text{reject}H_0|D, H_0) f(D)$$

Then, by definition of the α level of the permutation distribution,

$$P(\text{reject } H_0 | H_0) = \sum_D \alpha f(D) = \alpha$$

□

So we can conclude that, in particular, the false alarm rate of the η statistic permutation distribution, $f(X_\eta | D)$, is controlled.

Simulations In order to demonstrate the validity and applicability of the proposed test procedure, we conducted a variety of simulations. It will be demonstrated that the type I error rate (FA rate) can sufficiently be controlled given the proposed procedure. Furthermore, it will be demonstrated that the permutation distribution under H_0 indeed converges to a normal distribution for multiple different underlying data distributions. Additionally, one simulation will construct a specific case, where the aros test outperforms pairwise t-test and a one-way ANOVA. Lastly, it will be simulated how under specific conditions the statistical power of the aros test compares to pairwise t-tests and a one-way ANOVA. For every set of example data A , we generated 50 observation for three groups. The α -level was set to $\alpha = 0.05$ and f was defined as $f(x, y) = (x \cdot y)(x \cdot x)$ (see "The aros test explained"). All example simulations, as well as an implementation of the aros test are provided via <https://gitlab.com/TommyClausner/aros-test>.

Estimating the type I error (FA rate) The FA rate was estimated by 10000 simulations on uniformly distributed random data. A new data set was created for each simulation. For each test, 10000 permutations were performed to estimate the permutation distribution under H_0 . If the null hypothesis was rejected, the result of this simulation was set to 1 and to 0 otherwise. To obtain the final result, all individual test results were averaged, yielding a FA rate of 0.0503, which can be considered sufficiently close to the target of 0.05, as the deviation is only 0.6%. Figure 22 (A) depicts the cumulative average of the result vector. After around 1000 simulations, the FA rate converged to 0.05, with only minor fluctuations for the other 9000 simulations.

Demonstrating independence of sample data distributions A crucial step towards verifying the validity of the aros test is to demonstrate that the estimated permutation distribution under H_0 (the distribution D), indeed results in normal distribution for a variety of data distributions. As previously mentioned, A_i (each paired sample) and \mathbf{s} (the respective rank order shape obtained from the average over N observations in A), can be seen as independent random samples given H_0 . Therefore, the average of all $f(A_i, \mathbf{s})$ under H_0 represents an average of independent random values, which according to the central limit theorem, distributes normal.

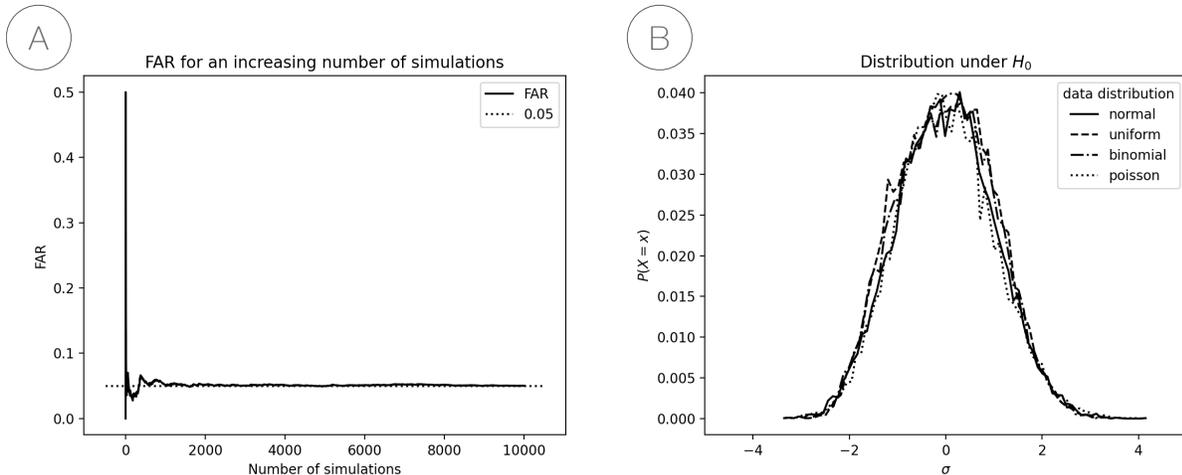


Figure 22: **A) False alarm rate under H_0 .** Each value of the continuous line represents the average of the result vector up to that respective simulation, whereas the dotted line represents the target value of 0.05 for the expected FA rate. Each value within the results vector could either be 0 (H_0 was accepted) or 1 (H_0 was falsely rejected). **B) Probability distribution under H_0 for a variety of input data distributions.** All distributions were standardized and transformed according to $Z = (D - \mu_D)/\sigma_D$. Afterwards the data was transformed into histogram data, where the count per bin was divided by the sum of all counts to obtain probability values for each point of the histogram. Histogram data was depicted as lines, rather than bars, for better readability.

For the purpose of demonstration, four different data distributions were used to estimate the permutation distribution under H_0 . Random samples from the following distributions (with stated parameters), were chosen: normal distributed data with $\mathcal{N}(0, 1)$, uniform distributed data with $\mathcal{U}(0, 1)$, binomial distributed data with $\mathcal{B}(3, 0.6)$ and Poisson distributed data with $\mathcal{P}(3)$. For each simulation 20000 permutations were used to estimate the permutation distribution under H_0 . The resulting distribution was standardized by subtracting the data to its mean and dividing it by the standard deviation. Afterwards, the data was transformed into histogram data with the number of bins determined by Sturges [387] or Freedman-Diaconis [388] rule. The approach yielding the higher number of bins was used, as is the standard for the `histogram(a, bins='auto')` function provided by numpy [389]. Furthermore, the count value for each bin was divided by all counts to obtain the probability for each point of the distribution. As Figure 22 (B) clearly shows, the resulting estimated permutation distribution approximates the same normal distribution irrespective of the underlying data distribution.

Simulated example In order to compare the aros test to paired sample t-tests and one-way ANOVA, we constructed a data set in the following way: A vector of $N = 50$ random standard normal distributed values was created, to which an offset of $[0.2, 0, 0.1]$ was added

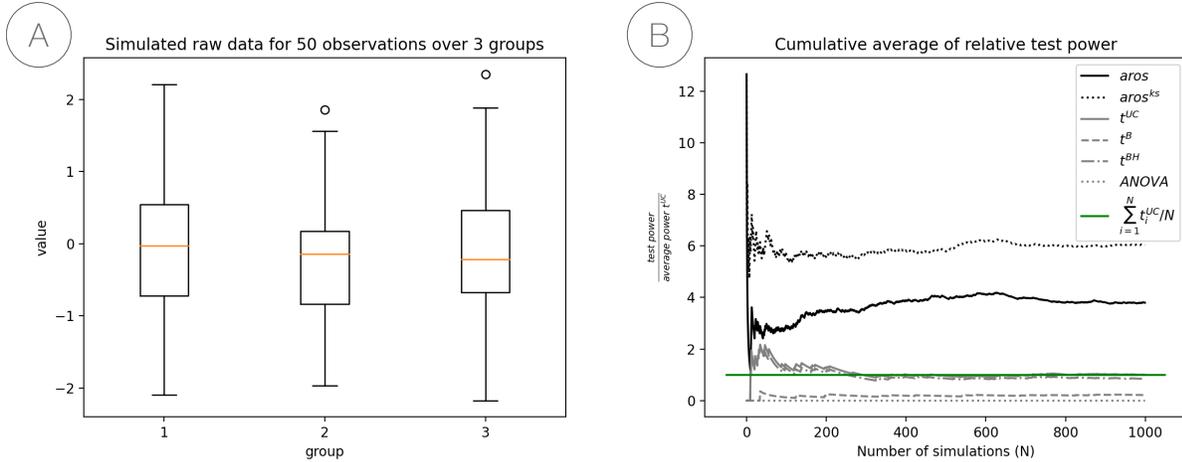


Figure 23: **A) Simulated raw data.** For each of the $k = 3$ groups, $N = 50$ observations were generated in the following way: N random samples were drawn from a normal distributions with parameters $\mathcal{N}(0, 1)$. To simulate the difference in means, the same data was used three times and shifted such that $\mu_1 = 0.2; \mu_2 = 0; \mu_3 = 0.1$. Additionally uniform random noise was added separately to each group, drawn from a distribution with parameters $\mathcal{U}(-0.5, 0.5)$. **B) Cumulative average of relative test power.** For each condition, the number of correctly rejected H_0 was counted and plotted cumulatively, relative to the number of simulations. This was done for $k = 3$ groups. Additionally, each value was divided by the final average of correctly rejected H_0 obtained from the uncorrected t-test scenario (t^{UC}). For each of the t-test scenarios, H_0 was counted as successfully rejected if all three of the pairwise comparison yielded a p -value lower than or equal to the critical α . This procedure was performed for the regular aros test (*aros*), a version of the aros test, where the shape was set to be known (*aros^{ks}*), pairwise t-test without correction for MCP (t^{UC}) and with correction for MCP using the Bonferroni (t^B) or the Bonferroni-Holm method (t^{BH}). Additionally a one way ANOVA for the simultaneous comparison of all three groups was performed (ANOVA). The green line indicates the final average of t^{UC} after 1000 iterations.

in order to simulate the difference between the group averages. Furthermore, uniform random noise was added to each group, drawn from a uniform distribution with parameters $\mathcal{U}(-0.5, 0.5)$. A box-plot of the simulated data for each group can be found in Figure 23 (A). The authors point out that the same random seed as for all other simulations was used and no particular choice towards tuning the result in a desirable way was made. However, the authors are aware of the fact that changing the random seed might indeed affect the clarity of the result. However, the aim of this particular simulation was to demonstrate that there exist data sets for which the aros test is particularly well suited. Furthermore, additional 1000 simulations were conducted, where multiple sets of data using the same parameters as in this simulation were constructed and the respective test power was compared.

In a first step the aros test was performed, followed by pairwise paired sample t-tests between the groups (1 vs. 2; 1 vs. 3; 2 vs. 3) and a one-way ANOVA over all three groups. The p -value for each test was obtained and in case of the aros test, the rank order shape as

well. Due to how the data was simulated, it is expected for the three averages to be related such that $\mu_2 < \mu_3 < \mu_1$. Respective p -values for each test can be found in Table 2. At a critical α -level of $\alpha = 0.05$, only the aros test and the t-test comparing groups 1 and 2 yielded a significant result. All other p -values were greater than the specified critical α -value. Additionally, from the aros test the rank order shape $\mathbf{s} = [3, 1, 2]$ could be obtained. Thus, the respective means can be ordered as $\mu_2 < \mu_3 < \mu_1$. This rank order could not have been obtained (given the data) using any other test.

Table 2: p -value of different statistical tests based on the data depicted in Figure 23 (A)

test	p -value
aros test	0.001
t-test 1 vs. 2	0.003
t-test 1 vs. 3	0.065
t-test 2 vs. 3	0.203
one-way ANOVA	0.618

Comparison of test power In order to approximate the relative test power given the scenario explained in "Simulated example", the procedure was repeated 1000 times (with similarly constructed, but newly generated data), where we recorded the number of correctly rejected H_0 for each test scenario. The following scenarios were included: aros test (*aros*), aros test with known (pre-determined) shape (*aros^{ks}*), pairwise t-tests with uncorrected p -values (t^{UC}), pairwise t-tests with Bonferroni (t^B) or Bonferroni-Holm (t^{BH}) correction and a one-way ANOVA (*ANOVA*) for simultaneous comparison of all groups. During each of the 1000 simulations, it was recorded whether H_0 was correctly rejected or not. For the t-test scenarios, H_0 was counted as correctly rejected, if p -values of all three tests (corrected or not), were lower than or equal to the critical α -level, which was set to $\alpha = 0.05$. For all other test tests, H_0 was counted as rejected, if the respective p -value was lower than or equal to α and the predicted shape corresponded to the presumed shape from which the data was constructed. This last condition was of course omitted from the ANOVA scenario. In a last step, the cumulative average of each condition was divided by the overall average of the uncorrected t-test scenario, which acted as a baseline. Figure 23 (B) depicts the cumulative, relative average for each condition. Thereby a change of the y -axis, e.g. by a factor of two, would indicate that the relative test power of the respective test was twice as high as the estimated test power of the uncorrected t-test after 1000 simulations.

None of the traditional methods was able to capture the respective difference between the three averages as good as the aros test. Note, that the result for the aros test with known

shape ($aros^{ks}$) needs to be taken with a grain of salt. Technically \mathbf{s} was not derived from an external source and must be considered "double dipping" in this scenario. Nevertheless, it was included to demonstrate the potential increase in statistical power for that case.

Discussion We propose a paired samples, non-parametric statistical test procedure on the basis of a permutation test, which aims to circumvent the MCP by combining the relationship between multiple group averages into a single statistical value. Via an auto-regressive procedure, paired sample data for multiple groups is related to the rank order shape of the group averages. As such, the aros test acts as a test for interval scaled data, where the test result needs to be interpreted on the basis of ordinal scaled ranks. It was demonstrated that the FA rate is controlled sufficiently by a permutation procedure, where the estimated permutation distribution D under H_0 can be described by a normal distribution with parameters $\mathcal{N}(\mathbb{E}[X_D], \sqrt{\mathbb{E}[(D - \mathbb{E}[X_D])^2]})$. A set of simulations has been conducted to verify the procedure. First, the FA rate was estimated over 10000 simulations, which yielded an estimated value of 0.0503. This value can be considered sufficiently close to the target value of 0.05. In order to save computation time, each simulation was based on 10000 permutations and $k = 3$ groups. Since it could be demonstrated that each $f(A_i, \mathbf{s})$ yields an independent random value, the authors do not assume the FA rate to change on the basis of the number of groups, nor by increasing the number of permutations per simulation. Furthermore, simulations on the basis of multiple data distributions showed that the estimated permutation distribution under H_0 distributes normal, as predicted. It needs to be pointed out that no particular reason for the choices of the data distributions can be brought forward. Instead - to the authors experience - the most common distributions have been chosen. Again, since each value in D stems from an average of independent random values, it is not expected that different choices for the data distributions would affect the normality of the estimated permutation distribution under H_0 . In a third simulation, the aros test was compared to the probably most common statistical tests: the t-test and ANOVA. Since the purpose of this paper was to demonstrate that conducting an aros test in some circumstances can be beneficial where t-tests or ANOVAs fail, the data was deliberately constructed such that the result would favor the aros test. However, no particular manipulation was applied to the data. Instead the data was constructed based on normal distributed data (required by t-test and ANOVA), where the difference in means was marginal compared to the noise that was added. This particular case (low signal-to-noise ratio) is one example where the aros test potentially outperforms classical tests. Nevertheless, the aros test has not been applied to a real data set in the current publication, which needs to be addressed in the future. However, one of the authors can confirm that the aros test has been applied successfully to a set of (yet unpublished) neuro-scientific data. In a last simulation, the relative power between the aros test and traditional tests has been assessed. For 1000 simulations, the true positive rate was compared between the tests. Thereby, the true positive rate of uncorrected t-tests served as

baseline and the result was computed relative to that average. Since the statistical power for tests on the null hypothesis is generally hard to assess and is strongly dependent on the data, the relative power to well established statistical tests might provide insight into how the aros test performs under specific circumstances. On average, the aros test performed more than three times as good as a uncorrected t-test, given the respective data. However, it needs to be pointed out that positive results for the pairwise t-tests were only counted if all three tests yielded a significant result. This procedure might be overly conservative in many circumstances. However, in this specific case, the distribution of the group averages was known beforehand and the goal was to determine the specific order in which those averages could be sorted. For such a statement to be made using t-tests, indeed three significant results would be required. Since not a single test using the one-way ANOVA yielded a significant result, the authors presume that it would not be suited for the respective data set. Relating within and between subject variance might have failed due to the large overall variability in the data that was much greater within than between the groups. Lastly, the aros test with pre-defined, known shape performed almost six times better than the baseline method. Since the present example would clearly be a case of "double dipping", this result should not be overstated. However, in a real world scenario, if the a specific shape would be expected (e.g. due to how the hypothesis was constructed or by previous research), this method potentially increases statistical power significantly.

In general, the biggest advantage of the aros test is its capability of allowing for a qualitative conclusion about the relationship between more than two group averages without prior knowledge of the underlying data distributions. Thereby, the test is relatively robust against noise. Depending on the choice for f , a significant result can be interpreted in multiple ways. Irrespective of f however, rejecting H_0 means that the paired observations $[1, \dots, N]$, are not fully independent and that the shape \mathbf{s} can be predicted by A with a probability $P(\mathbf{s}) > 1/k!$. This means that the labels of groups $[1, \dots, k]$ are not exchangeable since the distributions for $D_{[1, \dots, k]}$ are different. In other words, grouping the data into $[1, \dots, k]$ groups with labels $[1, \dots, k]$ is in fact meaningful and can be justified by the data. The respective choice for f enriches this finding with some additional information. If f was set to e.g. $f(x, y) = (x \cdot y)(x \cdot x)$, η indicates how many units of change in the data explain a single unit of change in the shape vector \mathbf{s} .

As mentioned before, the aros test is meant as an alternative for a variety of scenarios where more than two groups need to be compared, but classical statistical tests fail, either by violation of assumptions (e.g. violation of normality of the underlying data distributions) or if the MCP reduces statistical power unsatisfactorily. We could further demonstrate that the aros test potentially outperforms classical tests in low signal-to-noise ratio scenarios. Another positive aspect is the additional information that can be obtained by the rank order shape. An ANOVA, applied to multiple groups, can inform the user about whether there is

any difference across the tested groups, whereas the aros test additionally provides information about the qualitative relationship between the group averages. However, the relatively high statistical power to obtain an ordinal result from interval scaled data, circumventing the MCP, comes at a cost.

First and foremost the aros test does not provide any information about the quantitative difference between a set of means and can furthermore not be interpreted in such a way. Thus, it is impossible to obtain any effect size for the differences in means and hence should only be applied with this knowledge in mind. In some way the absolute difference between group averages is sacrificed in favor of an increased power for relative differences between the groups. Second, the result can only be interpreted in its entirety. This means no sub-comparisons between the relationship of means can be made and only the entire rank order profile as such can be interpreted. Note, that this is similar to the result of a cluster based permutation test [352]. This leads directly to a third caveat, that is the number of groups that can be meaningfully compared. While the interpretation of the rank order shape for three groups in most cases might be straight forward, it might not be for a large number of groups. In general, if a very large number of groups is compared and at least one dimension of the data is correlated, it might be advisable to choose a cluster permutation test [352]. Since it is possible to compute the aros test for an arbitrary large number of groups, the authors can only provide a rule of thumb, based on their experience in the fields of experimental Psychology and Neuroscience. The authors believe that the number of groups k to be compared, should be kept in a range where $3 \leq k \leq 7$. Values for k higher than that might be extremely difficult to interpret. Moreover, if H_0 was rejected, the obtained shape can only be interpreted if the group averages are indeed unique. This means that in order for the aros test to be interpretable, it needs to be ensured beforehand that each group average is numerically unique.

Lastly, the authors would like to point towards a variation of the aros test, which can be considered similarly valid as the standard procedure. However, if not applied carefully, this variation quickly leads to a circular analysis ("double dipping"). In some scenarios, the expected shape that is explained by the data might already be known. This can be the case if previous experiments allow for a justified prediction or in other cases the experimental hypothesis pre-determines the expected shape already. However, it needs to be emphasized that irrespective of the origin of the shape, it needs to be *independent* to not bias the estimation of the permutation distribution under H_0 . If such an independent shape could be derived from some external source other than the data, then this shape could be used to obtain η and computing D . Hence, one could test whether a *specific* rank order can be justified by the data with a probability higher than $1/k!$. Additionally, this approach can be combined with the classical procedure: If two independent data sets with k groups exist, then the classical approach could be applied to one of the data sets and - in case H_0 was

rejected - the obtained shape can be tested against the second data set.

Conclusion As demonstrated, the auto-regressive rank order similarity (aros) test can be considered an alternative test to circumvent the MCP for a small number of groups in a paired sample statistical test setting. While the ability to determine the magnitude of differences between group averages is lost, additional statistical power is gained to test the relationship between the raw data and the rank order of the group averages. Since the aros test is based on a permutation procedure to estimate the permutation distribution under H_0 , no assumptions about the distribution of the data are required other than exchangeability under H_0 . Furthermore, it has been demonstrated that the aros test controls the FA rate sufficiently. Since the aros test relates the magnitude of the group averages, without comparing them directly, it is exceptionally well suited for test scenarios, where the signal-to-noise ratio is low and the rank order of the means is of higher interest than the actual effect size.

Software developments for the EEG-fMRI experiment

In order to realize the EEG-fMRI experiment necessary to answer some of the research questions, many analyses strategies had to be custom designed, unified and brought into a format that can easily be shared and checked. While major methodological developments regarding the actual experiment are presented in the section which contains the paper in preparation ("Feature specific neuronal oscillations in cortical layers (in prep)"), this section focuses on developments that arose from data analysis, but are not in detail part of the main publication.

The main issue for laminar resolution simultaneous EEG-fMRI experiments is the relatively low SNR which imposes certain requirements on the experimental design and hence the analysis [390]. Not only is a relatively high number of participants required (52 in the present case), but furthermore analyses often involve manual intervention in order to maximize the signal quality. Particularly fMRI pre-processing steps (such as motion parameter estimation /realignment and functional to anatomical volume co-registration) turned out to be crucial processing steps with major influence on the final SNR [391]. A major break through could be achieved by employing accurate brain masks in order to limit the parameter estimation for motion parameter estimation to relevant areas [392]. However, this step required the manual correction of three-dimensional volume masks to almost voxel precision. Since the EEG-fMRI experiment was composed of a main task and extra blocks for pRF mapping, at least two masks per subject (one for each experiment) were required to be manually corrected. Thereby, existing tools, such as FreeSurfer's "freeview" or "tksurfer" [325], turned out to be quite cumbersome to use which slowed down the overall process. This led to the development of the "MRI volume masker 3000 TM" [329] which has been made publicly available (see below).

An additional challenge arose as a direct consequence of the experimental setup (EEG combined with high resolution fMRI) which pre-determined that the amount of data produced will be extensive. Furthermore, again in the spirit of open source and scientific transparency, as large as possible parts of the entire analysis pipeline were aimed to be more or less easy to replicate without too much in depth knowledge about the code. To solve both issues (large data and complex but easy to use code), a custom crafted meta script and configuration file system has been created that allows to pre-define crucial parameters for a respective analysis (e.g. motion parameter estimation) in a configuration file and returns an executable script where parameter settings are respected. Furthermore and if specified, each executable script can automatically be transferred to and executed at a compute cluster. Parameters with respect to wall time, memory, etc can as well be defined in the configuration file(s). In the current version the implementation relies on the compute infrastructure of the Donders Institute (DCCN), Nijmegen. Adjusting the system to a new environment with a different cluster job management however could easily be implemented by changing

the respective line in the main "wrapper" script. This system allows to run multiple analyses easily (by defining parameters and using the respective wrapper function), in parallel (shortening computation when dealing with huge data sets) and allows for sharing of exact processing scripts and functions that could be run independently. Both, the wrapper and analysis scripts have been made publicly available (see below).

Lastly, previous literature on simultaneous EEG-fMRI [238, 393] did not provide any methodological suggestion on how to statistically assess relative layer activation. Instead, a significant difference *between* α and γ activation *profiles* demonstrated feasibility but did show no way to map out *where* (in which layer) respective activation profiles differ and how. Traditional approaches (e.g. multiple t-tests with correction for multiple comparison) yield comparably low statistical power (compared to a e.g. a single t-test) and certain criteria (e.g. normality of the distribution) must be met. For those reason a custom made permutation test has been developed [353]. Instead of comparing a number of group means (or distributions), it is tested whether the rank order of the group means of the data can be explained better by the data than randomly shuffled data would explain *its* rank order obtained from *its* group means. If e.g. three groups a, b, c with means μ_a, μ_b, μ_c distribute such that $\mu_b > \mu_a > \mu_c$, a significant result would indicated that a, b, c can be meaningfully assigned to ranks [2, 3, 1]. If the null hypothesis cannot be rejected, no evidence for a given rank order has been found that could be explained by something other than a random order. The exact difference between group averages or effect sizes are unknown. Nonetheless, such test provides a useful tool for the respective layer analysis. A detailed description can be found in the section "Auto-regressive rank order similarity (aros) test (pre-print)". The code base, as well as all respective simulations used to verify the test are publicly accessible (<https://gitlab.com/TommyClausner/aros-test>) as well as an open access version of the publication (<https://www.biorxiv.org/content/10.1101/2022.06.15.496113v1>).

MRI Volume Masker 3000 TM has been developed to facilitate manual mask drawing on fMRI data [329]. It can be downloaded from:

<https://github.com/TommyClausner/MRI-Volume-Masker-3000-TM>

The interface and back-end has been developed in Python¹³ and must be executed in a corresponding environment. Required packages are installed automatically if the software is installed according to the specifications in the "readme". In a first step of the processing pipeline, the user indicates which file (.nii or .nii.gz) containing one or multiple fMRI volumes to load. A first brain mask is computed automatically using dipy [394]¹⁴. The main window depicts a large centered view of the middle horizontal slice (given RAS orientation¹⁵) flanked by a small sagittal and coronal view, where the respective position of the current central slice is indicated via a small line. Positive and negative parts of the pre-computed mask are overlaid with a semi-transparent layer in green or red respectively. The mask can be changed up to voxel precision with a closed loop selection. Using arrow keys it is possible to navigate quickly through the slices. Further, shortcuts allow for contrast changes to facilitate mask drawing in ambiguous contrast conditions, rotate between view perspectives and load and export masks. Since this tool was build solely for the purpose of fMRI mask drawing, it provides functionality to exactly and only do this, which in turn made mask drawing much easier than with commonly used alternatives from established software packages. Even though I am sure many researchers found their own solution to deal time-efficiently with cumbersome manual mask drawing, I believe that sharing solutions to even small problems helps the field as a whole to become more efficient.

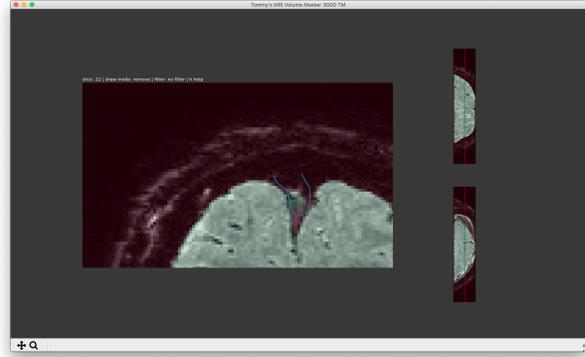


Figure 24: Example screenshot of MRI Volume Masker 3000 TM. It can be downloaded from <https://github.com/TommyClausner/MRI-Volume-Masker-3000-TM> including instructions for installation and usage.

¹³<https://www.python.org/>

¹⁴<https://dipy.org/>

¹⁵Typically fMRI data is four-dimensional (three spatial and one time dimension). RAS refers to the three spatial dimensions and how they related to real world topological landmarks. The three letters are meaningful in their arrangement, since they reflect the real world meaning of the respective axes XYZ. Furthermore, each letter corresponds to one of two choices which resolves which respective axis direction relative to a topological landmark is positive. Possible values are **R**: Right or **L**: Left; **A**: Anterior or **P**: Posterior; **S**: Superior or **I**: Inferior. Hence, RAS means that if the center of the coordinate system is place in the middle of the head, positive x-axis values are towards the subject's right, positive y-axis values towards the front and positive z-axis values towards the top of the head.

cluster script runner In the spirit of open science, sharing respective analysis scripts helps to verify published results, but furthermore offers the opportunity for researchers aiming to employ similar strategies to quickly implement novel methodological advancements. Thereby, the theoretical foundation as rolled out in a respective publication can be traced along the computational implementation. Lastly, encouraging the scientific community to self-evidently publish analysis scripts might inspire future projects but also incentivize to adhere to good code practice standards which improves overall "shareability". Of course, improved overall code quality and sharing rates should be compensated by visibility. Systems like "Zenodo" ¹⁶ allow to link software, data or other digital items to a digital object identifier (DOI), enabling or at least simplifying citability (see e.g. [DOI 10.5281/zenodo.7211758](https://doi.org/10.5281/zenodo.7211758)). In-

creased visibility (via citability) might increase overall compliance to a culture, where progress is not only shared in the domain of publications but furthermore accompanying efforts, such as analysis scripts or software tools are made publicly available too. However, sharing data and analysis scripts on a large scale, requires that certain standards are established. The Brain Imaging Data Structure (BIDS) standard [395] for instance has been proposed as a standard for organizing fMRI experiments and was later extended to included electro-physiological methods, such as EEG or MEG [396, 397]. Here, data collection for the EEG-fMRI experiment had started prior to the publication of BIDS extensions to EEG. It has finally been decided to craft a custom data organization scheme.

The scheme is fundamentally structured such that data and scripts are separated. Further, the data is structurally separated between *raw* and *processed* data, individually for each subject. Raw data includes the raw data, transformed into an easy to read file format (e.g. *.nii*). For each major step of the analysis, processed data is further structured into sub-directories. Within the scripts directory, multiple sub-directories include script templates, toolboxes, configuration files or analysis files that are shared across participants (e.g.

¹⁶<https://zenodo.org/>

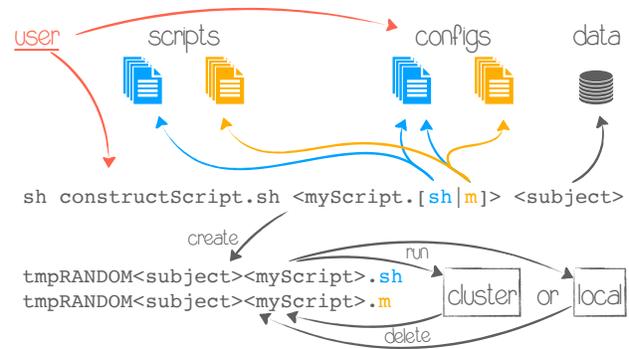


Figure 25: In order to speed up computation time, increase transparency with respect to analysis parameter settings, ease shareability and reconstruction of results, a custom made analysis management system has been implemented. For each analysis script (in the present case either bash or Matlab [348]) the script itself and a satellite configuration file (config), form the building blocks for the script generator from which a script is generated based on the provided template and parameters. This script is then executable on a local machine or compute cluster (provided a Linux environment). Once a respective computation is finished, each script deletes itself.

stimulus regressor for pRF mapping). Analyses are run, by calling a management script, including the target subject and target analysis, which composes a runnable script from a given template and corresponding configuration files. Afterwards, the script is submitted to the compute cluster, stored on drive or run locally. Analysis scripts are written in `bash` or `matlab`. This structure has been chosen, due to the large amounts of raw data that have been produced by the combine laminar resolution simultaneous EEG-fMRI experiment (see "Feature specific neuronal oscillations in cortical layers (in prep)"). In order to save computation time, individual subject processing steps (such as motion parameter estimation, co-registration, layer estimation, virtual channel estimation, etc.) can be run in parallel on the compute cluster (separate for e.g. each subject or ROI, etc). Furthermore, this system allows for a centralized control over the entire processing pipeline, which increases shareability. Even a novice to the code base necessary to replicate the presented result, could redo certain steps given certain parameters. In addition to that, systems like "Zenodo", require the underlying Github ¹⁷ or Gitlab ¹⁸ repository to be time stamped via a release. Thereby, the state of all files is fixed which enables locking the analysis pipeline to the state of publication. Hence, the state of each configuration file that comes with the shared code could be time stamped to submission or publication events.

In order to compute most of intermediate or final results, the user only has to specify respective compute scripts and participants. By default the management script composes the final runnable script and sends it to the compute cluster (at the moment based on the cluster architecture of the DCCN, Radboud University, Nijmegen ¹⁹). Parameter adjustments for each respective scripts can be done in a simple text editor. In most cases, the users interaction with the code base boils down to something like this:

```
for subject in 'seq 1 52'; do  
  sh constructScript.sh <script.[sh|m]> S$subject &  
done
```

Here, a respective script (e.g. `sh constructScript.sh eegPreprocessing.m S42`) is constructed for every subject (1 to 52) and sent for execution to the compute cluster. After the computation has been finished, the temporarily composed script deletes itself. Additional tools for live monitoring of jobs and cleaning up log files automatically, are provided as well. The analysis management tool can be downloaded from:

<https://github.com/TommyClausner/laminarfMRIv2>

¹⁷<https://github.com/>

¹⁸<https://about.gitlab.com/>

¹⁹<https://www.ru.nl/donders/>

However, as of today (January 15, 2023) the repository is still set to *private* mode. Conforming to the publication policies of collaborators, the repository will be set to *public* once the paper has been submitted (see current state at section "Feature specific neuronal oscillations in cortical layers (in prep)").

2.3 MEG experiments

As explained in the section "COVID 19 remark", MEG data collection has been affected largely by the global SARS-CoV-2 pandemic. In this section I will hence describe the progress of the current project(s) in a publication-like manner. However, since results are largely missing, theoretical and methodological considerations will be the center of the discussion. Some of the respective data, however, have been shared with collaborators and in return, preliminary results or results of particular aspects of those experiments that were of their particular interest have been kindly shared. It has been clearly indicated what the respective contributions have been and by whom they were performed.

Expected publication output

Ferez M., Clausner T., Lukacs J., Gbadoe M., Daligault S., Schwartz D., & Bonnefond M. Evidence of functional inhibition of high visual regions by α oscillations (in prep)

In addition to that, one first author publication is planned on laminar MEG with respect to attention modulation and predictability (Experiment 1, see page 123); one first author publication on laminar MEG with respect to higher / lower order visual region communication (Experiment 2, see page 130); and one second author publication - led by Maryam Mostafalu [1] (see below) - on MEG stimulus decoding with respect to attention and predictability.

Introduction

Some of the above research questions (see "Research question") could be addressed best using magnetoencephalography (MEG). Two MEG experiments have been conducted, specifically targeting hypotheses about the interaction of cortical oscillations on a laminar level. One of the first models relating spatially distributed α band oscillations to task-related, functionally relevant changes of cortical activity has been the gating by inhibition (GBI) hypothesis [224]. It has been thought that inhibitory α band activity reflects a gating mechanism that modulates the transfer of information by α amplitude and phase driven excitability changes in a cortical region. More precisely, specific changes of cortical excitability (reflected by α ERD and ERS) have been thought to enhance the processing of relevant and suppress the processing of irrelevant information [48, 168–170]. A possible hypothesis to explain those findings that translates to memory encoding performance too (lower α , better encoding), is that incoming signals are de-correlated in order to enhance information content [167]. In terms of Shannon entropy [145], the de-correlation linked to α ERD directly reflects increased information transferability. Hence, feedback directed changes in α band connectivity and power

might be explained by a mechanism lowering the threshold for signal transmission between regions. This would allow for feed-forward directed nested information flow in the γ band [91] (see "Communication via nested oscillations") due to a synchronization of excitability in source (lower order) and target (higher order) regions. Note, that recent findings challenge the view that α directly inhibits or enhances neuronal signalling [176] and instead imply a more information sampling related role [179, 180]. The idea would be that beside coherently activating communicating regions, the α (or θ) band phase gradient is used to sort information with respect to signal strength (strong signals are processed earlier in the inhibition cycle). Indeed, it could e.g. be shown that the respective α phase changes predict whether a respective target stimulus is missed [207, 210] which indirectly relates perceptual processing to rhythmic inhibition (high inhibition phase, target was missed). A similar mechanism has been proposed for the θ rhythm, where stimuli are "sorted" by saliency along the θ phase [180, 206] (see "Coding via nested oscillations"). Here, the role of α is expected to be similar, however, on the level of stimulus *features* (on a lower level e.g. edges or corners and on a higher level e.g. eyes or nose) rather than objects (on a lower level e.g. squares and on a higher level e.g. faces) [210]. Furthermore, α is possibly nested into θ [211].

Results from the EEG-fMRI experiment (see "Feature specific neuronal oscillations in cortical layers (in prep)") furthermore suggest multiple α band processes that fulfil distinct roles, such as attention or general activity modulation and feature selective activity changes. Thereby, a superficial layer α band process - possibly directed attention related [56] - could be successfully distinguished from a deep layer α process that not only exposed feature selectivity, but furthermore frequency specificity in the α band. The identification of two separate α processes has been predicted by recent theoretical frameworks [91]. Feedback related processing in deep layers has been reported from humans and animals [34, 52, 270, 272] and was (indirectly) found by the above mentioned EEG-fMRI experiment. Until now it has however not been shown - to the best of the author's knowledge - that feedback related connectivity changes are indeed linked to feature specific processes across different cortical layers in healthy human participants. To address this issue, laminar level anatomical resolution (to be able to identify cortical layers) and a high data signal sampling rate (order of *ms*) are required. Simultaneous EEG-fMRI could test for this indirectly, by relating the BOLD connectivity changes to respective spectral power changes [265]. However, a more direct estimation of frequency related changes could only be provided by laminar level MEG, for which feasibility has been demonstrated for sensori-motor related β burst mapping and the mapping of cortical oscillations in the visual cortex to a two layer model [220, 268]. Current developments in our lab bring even a 3 layer cortex layer into tangible proximity, by allowing the recording in supine (laying down) position (see "Methods"). However, the here presented MEG data has been collected before respective changes have been implemented (lab internal benchmarks using the obtained MEG data, computed by James Bonaiuto, indeed confirm the required spatial resolution). To investigate multiple aspects of oscillatory

activity with respect to the cortical architecture, inter-area communication, spatial and feature attention, as well as processing specificity, two laminar level MEG experiments, as well as multiple short fMRI experiments related to pRF mapping (determination of visual ROIs) and stimulus type localization (activation compared to baseline for a given stimulus type), have been conducted. MEG Experiment 1 has been specifically set up to test inter-area stimulus dependent frequency changes in early visual cortices across two cortical layers using simple grating stimuli. Experiment 2 in turn uses more complex (almost naturalistic) stimuli to trigger face or word processing areas (FFA and VWFA) and trace respective feature dependent feedback related α power changes and bottom up γ activity along the visual hierarchy.

Overview Experiment 1: spatial attention and predictability To investigate the link between α and attention, a lateral attention task was used. Two stimuli were presented on the screen, one of which should be attended, while the other one should be ignored. This procedure is known to elicit lateralized hemisphere specific α band modulation, where α power is lower in visual areas contra-lateral to the side of attention and higher ipsi-lateral [173] (Hypothesis 1), which has been further shown to be driven by frontal areas [251] (Hypothesis 2). Thereby, different levels of distractor inhibition are implemented by modifying a) the chance that the respective attention cue is valid (100% and 90%, named V100 or V90 respectively) - cue validity was shown to reduce attention α lateralization with decreased predictive value of the cue [398]; and b) the side of the cue indicating the switch of the side of attention - if the switch indicator lies on the unattended stimulus, the α lateralization should be reduced due to the reduced distracting "value" of the unattended stimulus (displays potential important information) [399]. If α is indeed related to attention specific feature sampling [91, 207, 210], then the lateralization of α should decrease the more likely task relevant information can be obtained from the actually unattended side. From the perspective of predictive coding, spatial predictions (attended target location) that are violated (in case of an invalid cue condition) would be expected to trigger respective responses in the γ band (prediction error), which are larger compared to the valid cue condition [219, 304].

Invalid stimulus conditions have hence been varied with respect to the cue location that indicates the switch. The cue to initiate the switch could either be presented on the former target *and* the distractor or on the distractor side *only*. Combined with the always valid V100 condition, this means that the distractor is a) meaningless to the task (V100), b) only useful to determine distractor orientation in case of a switch (V90b, for details see "Experiment 1: spatial attention and predictability" on page 123 for details on the nomenclature) or c) carries information about both: the switch *and* the potential new target orientation (V90s).

Stimulus and feature dependent α activity for different features (orientations) has been observed and related to deep cortical layer activity in primary visual areas of monkeys [52] (but see also "Feature specific neuronal oscillations in cortical layers (in prep)" on page 58). Using the here presented experiment, it is furthermore possible to test whether those findings can be replicated for (e.g. for the comparison between the left and right easy conditions) and whether indeed α feedback activity recruits stimulus feature dependent neuronal populations (Hypothesis 4). In humans, findings from laminar level fMRI indicate feedback feature related activity in deep layers as well [34]. However, as compared to laminar fMRI, laminar level MEG [220, 268] enables the investigation of feedback directed (attention related) activity with respect to frequency power changes. Again, feature specific neuronal populations are expected to be related to α power decreases in deep cortical layers (Hypothesis 6).

Stimuli were composed from simple gratings, known to elicit sufficient α and γ power changes in primary visual regions for this type of stimuli [189]. Due to the parallel lines each grating has a perceived orientation. The task was set up such that participants had to respond to the direction of the grating orientation (left or right) relative to the 45° implicit reference diagonal. For each subject it has been determined beforehand - in a staircase procedure - which deviations from the reference have been perceived as *easy* or *difficult*. From this, four possible stimulus combinations arise (orientation by difficulty).

On top of the variation of attention, stimulus predictability has been varied as well. The canonical microcircuits model [82] (see "Canonical microcircuits: cortical compute modules") links feature specific feedback directed predictions to deep connections and respective activity to the β bursts conveyed by α [219] (Hypothesis 5). In turn, according to the model, respective prediction errors would be transferred between regions via superficial layers and reflected in the γ band [82] (Hypothesis 7, Hypothesis 8 and Hypothesis 9), specifically transferred via α coherence [91] (Hypothesis 10). Here, predictability has been implemented by means of a Markov chain²⁰. Predictability was implemented as the fraction of trials for the four types that are preceded by a specific stimulus type. Each of the four stimulus types predicts the next stimulus type with 80%, 10% or $2 \times 5\%$ chance different for each type. Hence, with a probability of 80% the subject could predict the upcoming stimulus. Violations of those predictions - prediction errors - can be investigated with respect to γ in superficial layers, including inter-area coherence. It is expected that superficial feed-forward directed γ band activity is nested phase-dependent in α .

For a more detailed description of the applied methodology see "Methods" on page 122,

²⁰A Markov chain or Markov process is a series of events that are statistically linked to each other. Thereby, the current state predicts the next state with a certain probability, creating a sequence of probabilistic events that each are determined by the previous element.

especially "Experiment 1: spatial attention and predictability" on page 123 for experiment specific details.

Overview Experiment 2: feature attention and conflicting information In addition to investigating low and high frequency oscillations with respect to spatial feature attention and predictability, stimulus or even object specific oscillatory changes, including more long range connectivity changes (Hypothesis 2), are investigated in this experiment. Since higher order neuronal activity often encompasses entire networks [276], but here specific inter-area communication is targeted, functionally and anatomically separable (preferably with limited spatial extension) higher order regions need to be investigated. The fusiform face area (FFA) is known to respond relatively specifically to the presentation of faces [245, 281], whereas the visual word form area (VWFA) exposes increased activity for the shape of actual words relatively reliably, irrespective of e.g. casing [400] and is known to play a role in attention to those word shapes [401]. Both regions reflect anatomically different and functionally separable processes that can be exploited for investigating object or feature specific targeted α and θ conveyed long range connectivity and its relationship to forward encoded stimulus information, reflected by γ (Hypothesis 7 and Hypothesis 10) in more superficial and middle layers (Hypothesis 9). Hemifield specific decreases or increases in the α band, power contra- or ipsi-lateral to the attended stimulus [251], are hypothesized to extrapolate to feature attention processes as well (Hypothesis 3). This means that for rivalling features or information, respective target feature specific higher order areas experience a decrease in α power, whereas the rivalling other higher order region would experience an increase (Hypothesis 4). Both requirements (easy to identify, separate target areas and a functional rivalry between them) have been implemented in form of a Stroop task [278]. Higher order visual areas have been targeted by composing the stimuli from faces with a semi-transparent overlaid character string. Faces have been synthetically generated by averaging multiple publicly accessible images of public figures. Thereby, the assigned gender of the faces has been biased towards stereotypically perceived as more *female* or stereotypically perceived as more *male*. The four letter character strings that have been overlaid on top of the faces have been selected from publicly accessible name data bases for each of the two groups. Face and name could hence be either congruent or in-congruent. The elegance of those stimuli lies in the fact that low level primary visual regions (e.g. V1-4) are targeted similarly - to a large extent - by the compound stimulus, however higher order regions (FFA and VWFA) are targeted differently depending on the attended aspect of the compound image. Especially in the in-congruent condition a specific inhibition of the distracting information is expected (Hypothesis 1), related to MEG activity in the α [177] and β bands [254, 402], especially in deep [52], superficial (see "Feature specific neuronal oscillations in cortical layers (in prep)") or both layers [91, 272] (Hypothesis 6).

For a more detailed description of the applied methodology see "Methods" (below), es-

pecially "Experiment 2: feature attention and conflicting information" on page 130 for experiment specific details.

Methods

All MEG experiments are embedded in a series of experiments, including fMRI, EEG and MEG. Data has been recorded from five sessions for each subject where multiple researchers have been involved and recording time within one session has been shared, as well as e.g. anatomical data that was not experiment specific (see Figure 26 bottom). Both MEG experiments that are discussed here have been collected during recording sessions two to five (see details below). Since respective recording modalities are shared and only the stimulus and task varied, general aspects of the used methodology are presented for both experiments and only the different task modalities are described separately.

Subjects It has been attempted to record data from 40 participants that completed all experiment recording sessions. Due to five drop out participants ²¹, the initial set of participants has been extended to 45. From those 45 recorded participants, 40 completed all recording sessions (24 ♀) at an average ($\pm\sigma$) age of 26.4(3.19) years. Right-handedness was assured by the Edinburgh Handedness Inventory [403] and participants had either normal or corrected to normal vision. Both experiments have been approved by the local ethics committee. Formal consent was collected and participants underwent a medical screening prior to the experiment, where general medical condition and MEG compatibility was assessed. Furthermore, participants were asked to fill a set of questionnaires, asking general life style related questions, such as the overall sleep quantity and quality, as well as alcohol and / or other drug consumption prior to the date of recording.

Data Acquisition MRI and fMRI data has been recorded using a Siemens Sonata 3T. Using a T1 and a T2 weighted MP RAGE sequence [310] with a resolution of 1 *mm* isometric voxel size ($256 \times 256 \times 256$ voxel), anatomical images have been acquired. Those images have been furthermore used to create individual head-casts that have been manufactured for each subject in order to reduce between session co-registration errors and within session movement. In particular, the scalp surface was extracted from MR images for each subject and used as a template for the inner shape of the head-cast. The outer shape resulted from the MEG dewar's inner shape [266, 268]. Each individual head-cast has been built by Denis Schwartz and Sébastien Daligault from the cermep imaging center ²². In a first step, the 3D surface meshes obtained from individual MR images have been pre-processed

²¹The recording was massively impacted and delayed by the COVID 19 pandemic which made it impossible for some participants to adhere to the initially agreed time schedule that has been rolled out beforehand.

²²<https://www.cermep.fr/>

in Rhinoceros 3D ²³ in order to make them "water tight" for 3D printing and exported to .stl. Afterwards, individual head-shapes have been 3D printed using a "Raise 3D N2 plus" printer ²⁴. Hereafter, replica heads have been placed inside a negative replica of the MEG's dewar. The space between both surfaces has been filled with polyurethane foam ²⁵. After hardening (and minor individual adjustments for the subject's comfort), the head-cast precisely fills the space between the subject's head and the MEG dewar. This greatly increases inter-session repeatability of subject positioning, as well as it reduces motion of the subject during the experiment below the 1 *mm* mark [266, 268]. In addition to anatomical scans, fMRI data of a more or less standard pRF mapping experiment has been recorded [72] that will not be rolled out here in detail. Most importantly, this experiment was conducted to obtain primary visual areas for each subject (V1-4) as well as - in a separate task - localize certain task specific areas, such as FFA.

MEG data has been recorded during each session with the same individual head-cast for each subject using a CTF Omega 275 channel system ²⁶ in a magnetically shielded room at a sampling rate of 1.2 *kHz*. Three fiducial marker coils have been embedded into the surface of the head-casts at nasion, left and right ear canals, in order to measure the subject's relative head motion. In addition to that, eye tracking data has been recorded using an EyeLink 1000+ ²⁷ at a sampling rate of 1 *kHz*. Stimuli have been projected to a screen of ≈ 50 *cm* distance to the subject's eyes. A two button button-box was used for the participants to indicate their responses with their right hand's index and middle finger.

Experiment 1: spatial attention and predictability Data recordings for each subject in this experiment has been distributed over four MEG recording sessions. This was done due to the excessive amount of trials that have been collected across the entire experiment. Since the respective task consumes a lot of attentional resources, too long recording sessions would have possibly hampered the subject's performance. Additionally, ethical concerns regarding a multi-hour recording session could be brought forward as well. Due to the amount of development that went into the individual head-cast construction [220, 268] for each subject, inter-session positioning errors have been below or around 1 *mm*.

Stimuli were composed from gratings () with 12 cycles across and 3.5° visual angle in diameter (higher spatial frequency than what is shown here), similar to what has been

²³<https://www.rhino3d.com/>

²⁴<https://www.raise3d.com/>

²⁵Flex Foam-iT!III - <https://www.smooth-on.com>

²⁶<https://www.ctf.com/>

²⁷<https://www.sr-research.com/eyelink-1000-plus/>

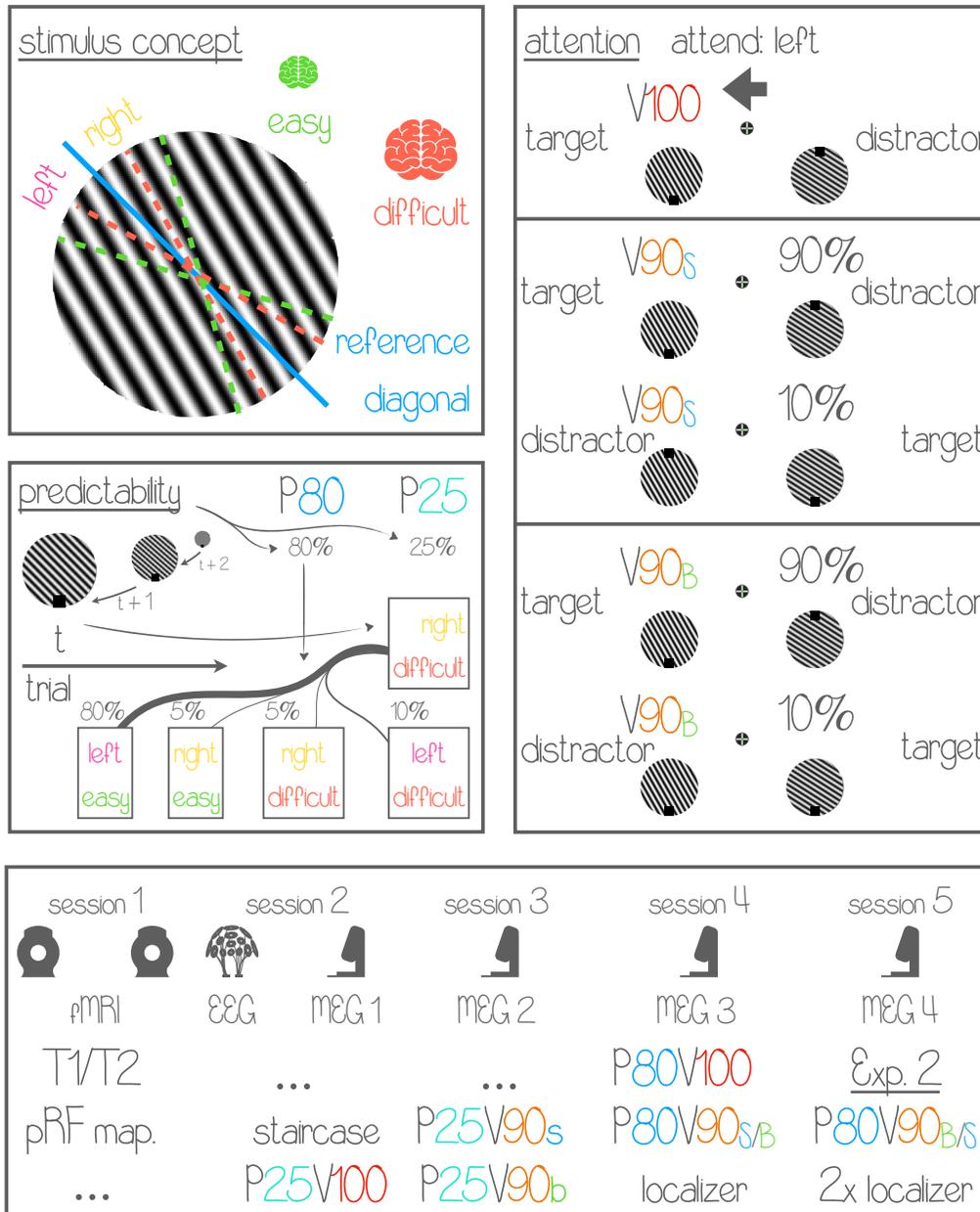


Figure 26: **Stimuli, attention conditions, probability conditions** and overall **recording schedule**. Top left: example stimulus. Deviation from the diagonal to the left or right has been varied for two levels of difficulty. Right: Variation of cue validity (see page 126). Middle left: Variation of stimulus predictability (either not predictable: P25 or predictable with 80% probability: P80). See Figure 27 and page 126. Bottom: Overall recording schedule. Each subject was recorded in 5 sessions. In each session, experiments, experimental conditions and recording methodology varied.

used in the EEG-fMRI experiment (see "Feature specific neuronal oscillations in cortical layers (in prep)"). A central fixation target (), designed to maximize fixation performance [316] was placed at the central screen location. The task of the participants was to determine the orientation of the parallel grating lines with respect to the imaginary diagonal. The reference diagonal could either be left (\) or right (/) with a $\pm 45^\circ$ angle relative to the vertical axis. For each subject the reference diagonal would always remain the same over all experimental sessions, but was counterbalanced across participants. Deviations from the diagonal could be achieved by clock or counterclockwise rotation of the grating stimulus. A stimulus rotation of an angle *greater* than the diagonal would be called a *left* stimulus (for left diagonal: stimulus rotated closer to horizontal axis; for right diagonal: stimulus rotated closer to vertical axis), whereas a stimulus rotation of an angle *less* than the diagonal would be called *right* (for left diagonal: stimulus rotated closer to vertical axis). With their right hand index and middle finger participants indicated for each trial whether the stimulus orientation deviates to the left or right with respect to the assigned diagonal (index finger: left). The diagonal itself is not be visible. See Figure 26 top left for an illustration.

Task and stimulus composition varied depending on the respective (sub-) experiment. The first MEG experiment has originally been designed by a post doctoral researcher has left the lab unexpectedly before the recordings fully started. Since respective recording sessions had already been scheduled, the task was slightly modified to serve as a staircase ²⁸ experiment for this *Experiment 1*. During the staircase procedure the deviation from the respective diagonal has been varied between $\pm 1^\circ$ and $\pm 44^\circ$ relative to the diagonal. Difficulty was assessed by the distance of a stimulus rotation to the diagonal. Furthermore, since the original experiment was planned as a spatial attention task, a stimulus could either be far away from the center of the screen or close to it with a different visual angle of $xy_{close} = 3^\circ$ or $xy_{far} = 4.4142^\circ$. Stimuli have been shifted towards the bottom right of the screen with the respective distances above. For the staircase procedure, stimulus presentation was uni-lateral, however central fixation was instructed and assured by eye tracking. For both, close and far stimuli, the deviation from the central diagonal that yielded 80% correct responses has been chosen as a reference for later experiments. Thereby, the difference to the diagonal at

²⁸The staircase procedure is a method to map certain psychometric values. Thereby, features of a stimulus (e.g. loudness) are modified systematically, such that sensory thresholds can be determined. For example, participants could be presented with a variety of tones of different frequencies and amplitude, to map the range of hearing for each frequency band. Instead of collecting many examples for every frequency, the difficulty could be adjusted based on previous responses. For the auditory example, one could present a specific tone at a specific frequency and vary amplitude or frequency, such that the tone is perceived less loud. Once the tone is not heard anymore (e.g. no response), subjective loudness is increased again, until a threshold is found, where a certain percentage of correct responses is reported on average. This adaptive procedure greatly facilitates the parameter estimation for certain stimulus parameters, such as difficulty.

80% correctness for close stimuli is typically smaller, because the stimulus is closer to central fixation and hence the deviation can be spotted more easily. Therefore, angles derived from the *far* condition (that are typically greater) have been used as *easy* and respective *close* condition trials (typically smaller angle) were used as reference for the *difficult* condition in the later experiment. For each subject, individual easy and difficult stimulus orientations have been obtained that way. Since the staircase procedure needed to be performed prior to the main experiment, all staircase recordings have been conducted during recording session 2 (see Figure 26).

Main experiment Two grating stimuli have been presented lateral (left and right) on the screen, shifted towards the bottom ($xy \pm 3^\circ$ visual angle from the center). Subjects are asked to report the relative orientation of the grating on the *attended* side (left or right) and ignore the respective other stimulus (indicated before each trial). During each experimental block, the respective target attention side would remain, but would change counterbalanced (50 : 50) between blocks. The cue could either be valid or invalid. Depending on the condition, the target stimulus is displayed on the respective other side of the screen with a chance of either 0% or 10% (odd ball). The condition where the cue is valid in 100% of all trials is called "V100" (100% valid) or "V90", when the probability that the cue is valid is only 90%. A respective change of target side is indicated on the stimuli themselves. For the main experiments, each stimulus was modified to contain a small black square either on the top or bottom of the stimulus. On the attended side the black square would usually

be at the bottom (), whereas on the distractor the black square would be at the top (). In reality, black squares were smaller relative to the stimulus than depicted here. A change of attention side is indicated by the black square of the distractor stimulus moving to the bottom. However, this can happen in two ways: in the *s* or switch condition, the black

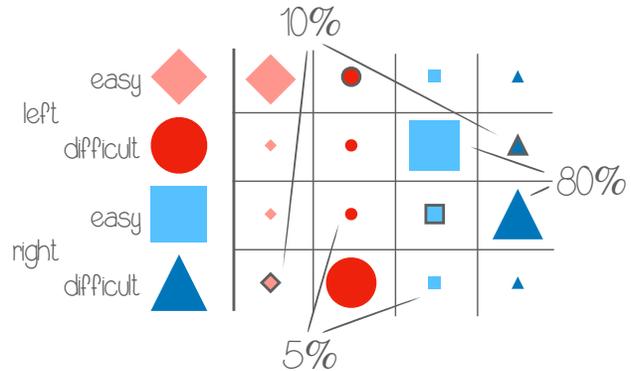


Figure 27: **Stimulus predictability** Each leading stimulus predicts a trailing stimulus (the next stimulus) with a certain probability. Thereby, one stimulus is predicted with 80%, one with 10% and two with 5% probability. The two highest probability stimuli are thereby chosen, such that that the respective button press can be predicted with 90% probability. For each subject such a table has been computed. In order to counterbalance respective stimuli, leading stimuli for respective trailing stimuli have been rotated (in the figure this would correspond to switching the words at the symbols). Hence, four possible combinations arise (one of which is the presented example) that are assigned counterbalanced to respective participants (10×4 sets of participants). See Figure 26 for an overview of how this is embedded into the grand experimental scheme.

square is moving to the bottom at the distractor side (making it the new target), which is accompanied by a switch of the black square on the formerly attended side from bottom to top. In a second condition, only the black square on the distractor side moves to the bottom (making it the new target) and the black square on the formerly attended side remains at the bottom. This condition is called *b* for "bottom". Note, that in the first condition there is a cue on the originally attended target, which indicates the switch, whereas in the second case this indication is not present on the target side, requiring the subject to "monitor" the unattended side as well. Hence, the experiment was conducted in three variations of cue validity: 100% validity (**V100**) and 90% validity, where either the switch of attention is indicated by a switch of the black boxes (**V90s**) or by the black boxes of both stimuli moving to (or staying at) the bottom (**V90b**).

Each of those three experimental conditions contained 1600 + 320 trials. The separation comes from an additional variation that has been added: stimulus predictability. Since stimuli were composed based on whether they are left or right oriented and easy or difficult (difference to the reference diagonal is large or small), they can be sorted into four groups: left and easy (**Le**), right and easy (**Re**), left and difficult (**Ld**) and right and difficult (**Rd**). Predictability has been introduced by adding a certain probability at which a certain stimulus from the four groups is preceded by another stimulus of those four groups. This could e.g. be an 80% chance that stimulus Le is followed by Rd. As a control, a set of 320 trials (80 trials per block) for each of the three attention conditions has been collected, where the predictability for each given stimulus was 25% (chance level; no predictability). Additionally, 1600 trials for each of the three attention conditions have been collected *with* varying stimulus predictability. Thereby, stimuli are generated by a Markov chain, such that each preceding stimulus predicts the next stimulus with a certain probability. Subjects have not been informed about the stimulus predictability. Each stimulus could predict another stimulus with either 80%, 10% or $2 \times 5\%$ certainty. Thereby, the stimuli that are predicted with 80% certainty and 10% certainty form a unit, such that the respective response button for both is the same (different difficulty). This means that a certain response can be predicted from any given stimulus with 90% probability. The remaining 10% are distributed evenly for the remaining two stimuli. This means that a certain stimulus predicts a certain other stimulus with 80% chance, which already determines the 10% stimulus (same orientation but different difficulty) and the remaining two (each 5%). To counterbalance across "leading stimuli" that predict their respective "trailing" stimulus, probabilities for each stimulus to predict another stimulus with 80% have been rotated. For example, the following leading-trailing stimulus pairs could predict one another with 80% probability: Le \rightarrow Le, Ld \rightarrow Re, Re \rightarrow Rd and Rd \rightarrow Ld. From there, 10% probability stimuli can be derived: Le \rightarrow Ld, Ld \rightarrow Rd, Re \rightarrow Re and Rd \rightarrow Le, which further determines the 5% stimuli. Further counterbalancing is applied by not only varying the trailing but furthermore the leading stimulus. For each stimulus (e.g. Le) the respective 80% trailing stimulus could be Le, Ld,

Re and Rd. This leads to four sets of four predicting stimuli. One of those sets is used for one subject throughout the entire experiment and all four sets are balanced across all 40 participants. See Figure 27 and Figure 26 for a graphical representation. Predictability conditions are called either "P25" (320 trials for each attention condition) if the stimulus was predictable with a chance of 25% (chance level) or "P80" for 80% predictability (1600 trials for each attention condition). All experimental P25 blocks have been recorded either during session 2 (P25V100), session 3 (P25V90s and P25V90b), session 4 (P80V100 and P80V90s or P80V90b) or session 5 (P80V90b or P80V90s). If P80V90b was recorded during recording session 4, then P80V90s was recorded during session 5 and vice versa. Whether P80V90s or P80V90b was recorded during session 4 for was counterbalanced across participants.

Experimental blocks and trial lengths have been designed with maximum trial count per amount of time in mind. However, given certain requirements imposed by the MEG recording software only ≈ 10 *min* segments of data could be recorded. To keep trial counts stable across blocks, P25 trials have been divided into chunks of 80 trials per block, whereas 100 P80 trials formed a block in that condition. Before the start of a new block, a 5 s baseline recording period with only central fixation has been recorded, where the subject was asked to avoid blinking. After that, the subject would receive instruction about the attention condition (attend left or right) and whether the attention cue is *always* (V100) or *mostly* (V90) valid. Each trial of the experimental block that followed hereafter was composed as the following set of events: A baseline period of 1.2 s (up to ± 0.2 s) jitter preludes the stimulus presentation (fixation indicator only). Thereafter, the stimulus is presented (two lateral gratings and central fixation) and the subject is asked to indicate the respective deviation of orientation of the target stimulus as fast (but reliable) as possible. With the button press the trial ends and the 1.2 s baseline period preceding the next trial begins. Subjects have been instructed to blink at the same time of the response (button press). Once the block has ended, the subject receives feedback in form of a percentage of correct answers during that block and additional verbal feedback, such as "Perfect!", "Very good", "Good" or "You can do better", depending on the overall performance. Within each recording session, experimental blocks have been shuffled with some minor exceptions. During recording session 4, where two P80 conditions (2×1600 trials) are recorded one block lasts for 100 trials. In order to facilitate implicit learning of predictability, the first two blocks of this session have always been V100 blocks. During those blocks, the subject does not need to focus on whether the side of attention might switch and it has hence been hypothesized that this helps to pick up the stimulus predictors. After those initial V100 trials, 24 randomly shuffled blocks of V90 and V100 are presented intermixed (16 V90 and 8 V100). The last 4 blocks have again been fixed to V100, since those blocks could be removed more easily if the attention span of the subject did not last long enough and they wanted to abort the experiment. However, this happened only two times over all participants.

Experimental procedure Each subject was recorded multiple times across multiple recording sessions and experiments. This part will only focus on the experimental procedure of recording sessions 2 to 5, where data for the presented work have been collected. On arrival of the subject, informed consent and a COVID 19 related questionnaire has been collected, as well as additional questionnaires related to sleeping behavior and general life style. During session 2 (where the EEG recording took place: steady state visual evoked potential (SSVEP) -task; not related to the two experiments here), the EEG cap was now fitted and electrode gel was applied to ensure proper conductivity. Afterwards, a set of photographs from the subject wearing the EEG cap was collected for later photogrammetry based 3D head reconstruction and electrode co-registration [313]. The EEG recording has been performed hereafter inside the magnetically shielded MEG scanner room. Once the EEG experiment had ended, the participants could voluntarily rest for a couple of minutes, before they received instructions about the upcoming MEG experiment and a practice block with increased odd-ball trial probability (no stimulus predictability) outside the scanner. When the subject felt comfortable with the task, the head-cast was fitted, reference coils were added and the subject entered the scanner room, where it performed the staircase and P25V100 task. During recording session 3, the overall procedure remained similar, except that the first experiment (a motor related task) was a MEG experiment conducted by yet another researcher. Between this experiment and the 2×4 intermixed P25V90s and P25V90b block instructions have been refreshed and another practice block was performed outside the scanner room. Session 4 was solely occupied by the above described procedure of intermixed recording of 2×16 blocks of P80V100 and P80V90s or P80V90b after the subject had again been instructed and had received practice outside the scanner. Lastly in session 5, either 16 blocks of P80V90b or P80V90s have been recorded after recording the data for Experiment 2 (below), depending on which version of the task was presented in session 4. After the main experiment, a localizer experiment was conducted, consisting of 80 trials of the main experiment, where the task was modified to a simple color change detection of the central fixation indicator.

All respective scripts to set up and run the experiment are implemented in Python using PsychoPy ²⁹ and will be available after submission of the publication at:

<https://gitlab.com/TommyClausner/meglyon> and

<https://gitlab.com/TommyClausner/meglyon-experiment-runner>.

Due to the delayed recording schedule caused by the pandemic, the respective data of this experiment has not yet been fully analyzed. Preliminary results and remaining work are discussed in section "Experiment 1: spatial attention and predictability" on page 134.

²⁹<https://www.psychopy.org/>

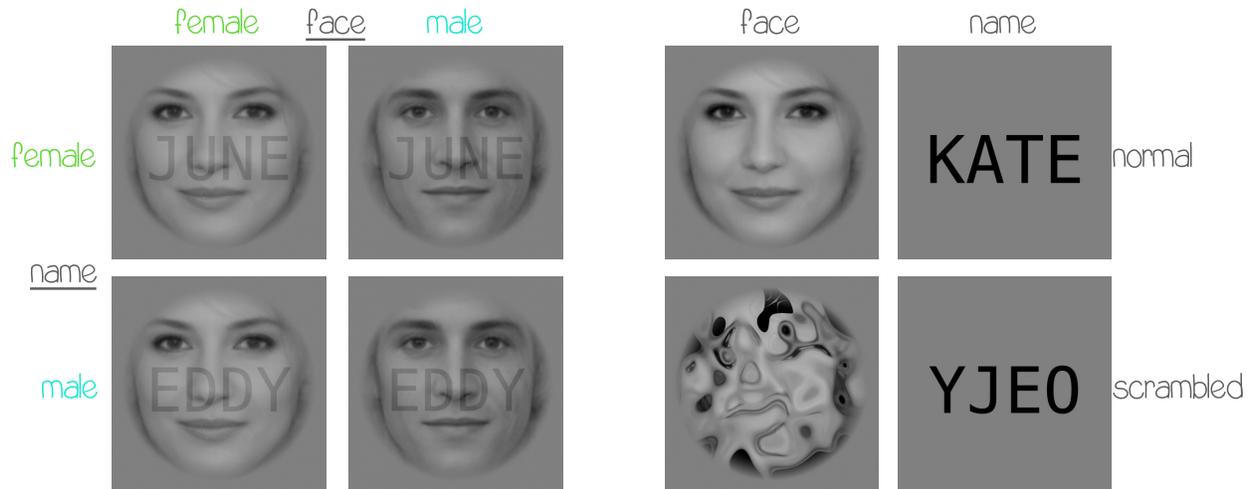
Experiment 2: feature attention and conflicting information Experiment two targets specific compute pathways to FFA and VWFA by using multi dimensional stimuli. In order to minimize the influence of the stimulus material on the response, stimuli were composed as semi-transparent overlays of each other. To probe the system and trace respective higher order stimulus dependent α feedback, conflicting information stimuli, similar to Stroop stimuli [278], have been used. However, as compared to words printed in an (in-) congruent font color, we aimed for more naturalistic and easy to trace stimuli, namely faces, which are known to elicit FFA activity [404]. Those faces were centrally overlaid with semi-transparent four character all capital letter words. Faces could expose stereotypically assigned male or female features and names have been selected from publicly available data bases, where a gender was asserted already ³⁰.

Stimuli were created using the VGGFace2 data [405], a multi-million labelled image data set containing faces. This data set consists of photographs of people, most of whom are public figures. In an attempt to avoid assertion of a gender solely based on the looks and external cues of unknown people’s faces and since well recognizable public figures necessarily needed to be excluded (to avoid some stimuli being recognizable), respective underlying faces have been artificially generated. Before face stimuli could be synthesized, a pre-selection was made in order to ensure that faces that compose the later average faces are photographed from a more or less frontal perspective. Furthermore, the sub-selection was based on whether the depicted person evidently self-asserts to a certain gender or is known to do so. For each of the two categories (female or male), 60 images have been selected. By using a pre-trained network provided by Dlib ³¹, 63 facial landmarks (such as eye corners) are automatically extracted. Those landmarks can be used to morph two (or more) faces towards each other. From the set of 60 available images, ten have been chosen randomly and the average face was computed. After that, a name from a set of 2×50 names is chosen to be overlaid with the face, depending on the condition (congruent / in-congruent).

After that and depending on the condition (congruent / in-congruent) a name from a set of 2×50 is chosen to be overlaid with the face. In total there are four possible conditions: $[\text{♀}_{face}, \text{♀}_{name}]$, $[\text{♀}_{face}, \text{♂}_{name}]$, $[\text{♂}_{face}, \text{♀}_{name}]$, $[\text{♂}_{face}, \text{♂}_{name}]$. Each stimulus has been overlaid with a similar fixation indicator as used in Experiment 1 () [316]. See also Figure 28 for a depiction of example stimuli and an overview of the recording procedure.

³⁰Note that the author of this thesis does not support or believes in any binary gender or sex system. Only in order to obtain *two* stimulus conditions we chose the respective reference frame. Subjects have been briefed about the respective setting and that they could at any time drop out of the experiment if they feel uncomfortable with the task.

³¹<http://dlib.net/>



main experiment → two localizers

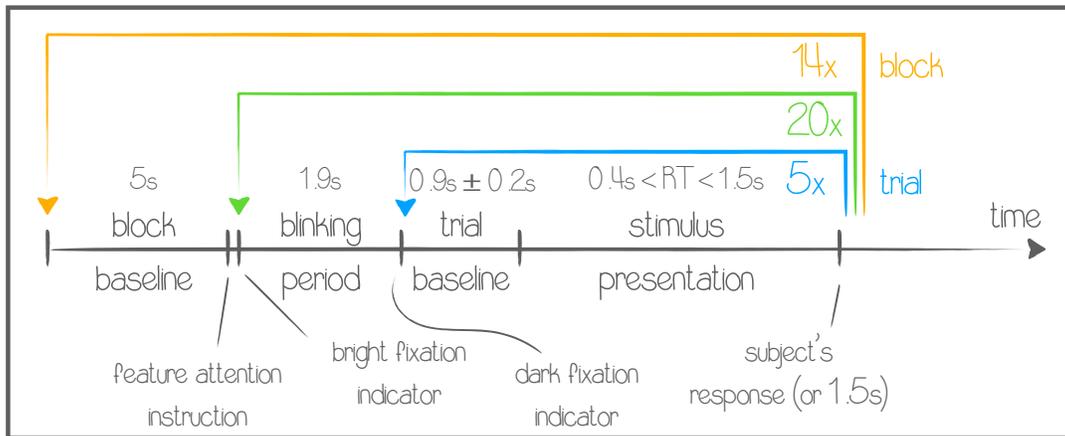


Figure 28: **Stimuli and experimental procedure** The top of the figure depicts example stimuli for the different congruence conditions. Faces have been generated from manually selected examples contained in the VGGFace2 data set [405]. If either a person self-evidently self-asserted to a certain gender or is known to do so (the data set contains mostly photographs of famous people) and the photograph was taken with a more or less frontal perspective, it had to be selected. From a set of 60 images for each of the stereotypically assigned gender groups, ten have been selected randomly and averaged using 63 facial landmarks. Landmarks were detected using a pre-trained ANN. Four-character names for each category have been selected from publicly accessible name data bases and randomly combined with generated faces. To localize FFA and VWFA respectively, two localizer experiments (one for names and one for faces) have been conducted using scrambled faces (using diffeomorphic transformations [406]) or non-words. The bottom half depicts the recording scheme for the main experiment. For each congruence condition, 350 have been recorded (1400 total). The experiment was split into 14 blocks with 100 trials each. A 5 s baseline was recorded, prior to the attention instructions of each block. Afterwards, a 1.9 s period was allocated to allow the subject to blink. After that, 5 trials are recorded ($0.9 \text{ s} \pm 0.2 \text{ s}$ baseline and maximum 1.5 s stimulus). During each trial, participants are asked to indicate the respective stereotypically assigned gender for either the name or the face. The trial ended with the subject's response, but lasted at least 0.4 s. See also Figure 26.

Experiment and trials have been again set up to maximize trial count. Thereby, the trial count of Experiment 2 has been weighted against the trial count of Experiment 1, both competing for experimental recording time in recording session 5. Since Experiment 1 contains extremely unlikely conditions, the experimental schedule during session 5 had been shifted towards Experiment 1. To maximize trial count in the remaining recording time, trials were presented in streams of five: Before the onset of a set of five trials, participants received 1.9 s explicitly allocated for eye blinking, which has been visually indicated by the fixation indicator increasing in brightness. After that, the fixation indicator would turn back to medium-gray (indicating the subject to avoid blinking from now on) and a baseline period of $0.9 \text{ s} \pm 0.2 \text{ s}$ is recorded. Thereafter, the stimulus is presented for at least 0.4 s and for a maximum period of 1.5 s but otherwise ended at the subject's response time. Baseline period and stimulus presentation alternate five times, before the blink indicator turns bright again, which indicates to the subject that it can blink now. Before each experimental recording block (100 trials), a 5 s baseline period of resting state MEG activity was recorded where the subject was asked to avoid blinking. For each of the four possible stimulus combinations 350 trials have been recorded, summing to 1400 trials in total. Those have been divided into 14×100 trials per block. During each trial, participants were asked to indicate the stereotypically assigned gender of either the name or the face via button press with index and middle finger of the right hand. Button mappings with respect to male or female have been counterbalanced across experimental blocks. Each time a new block starts, instructions on the current button mapping are displayed. See also Figure 28 bottom for a depiction of the general experimental recording procedure.

Experimental procedure After the subject's arrival, informed consent and a COVID 19 related questionnaire has been collected as well as additional questionnaires related to sleeping behavior and general life style. Afterwards, the subject received instructions on the task, the head-cast was fitted and reference electrodes were attached before it went into the scanner room. Before the start of each block, it was indicated, whether the subject was asked to report the stereotypically assigned gender identity of either the face or the name, while preserving central fixation. Furthermore, the respective button mapping ($[left, right] \rightarrow [female, male]$) changed counterbalanced across blocks, which was indicated prior to the onset of data collection for each block. After recordings of the main experiment were completed, two additional localizer tasks (160 trials each) were performed. In the face localizer task either a regular face that has been used throughout the face experiment or a smoothly scrambled version of that face, transformed using diffeomorphic transformations [406] (see example in Figure 28) was shown. Diffeomorphic transformed face stimuli have been used in order to keep image attributes (brightness contrast, etc.) and local correlations more or less similar, while removing the "facial content". This way, specifically the FFA can be targeted by contrasting normal and scrambled faces. For each type of stimulus (normal

or scrambled), 80 trials are shown randomly ordered in one block. The second localizer functions similarly to the first one, only that instead of faces and scrambled faces, respective names and four letter pseudo-words are shown. Again, 80 trials for each condition are used to localize VWFA.

All respective scripts to set up and run the experiment as well as the scripts used to create the compound stimuli from average faces are implemented in Python and will be available after submission of the publication at:

<https://gitlab.com/TommyClausner/meglyon> and

<https://gitlab.com/TommyClausner/meglyon-experiment-runner>.

Due to the delay in recording schedule caused by the pandemic, the respective data of this experiment has not yet been fully analyzed. In section "Experiment 2: feature attention and conflicting information" on page 143, preliminary results and remaining work are discussed.

Open research and discussion

As explained in the respective section ("COVID 19 remark"), data recordings for both MEG experiments suffered from substantial delays. Hence, only the planning, task development, implementation, piloting and recording of the data could be completed for both experiments. However, other members of our lab already analyzed or pre-analyzed certain aspects of the data which concerns Experiment 1 (see "Experiment 1: spatial attention and predictability") and Experiment 2 (see "Experiment 2: feature attention and conflicting information"). Most notably, Oussama Abdoun ³² analyzed the behavioral data and Maryam Mostafalu [1] analyzed a part of the sensor level data for Experiment 1; and Maxime Ferez analyzed data of Experiment 2 in sensor and (classical) source space (no laminar level source reconstruction yet) for his PhD thesis [407]. With their kind permission I will reproduce some of those results in order to justify future analysis steps, confirm first predictions or obtain intermediate result verification in form of sanity checks. I will meticulously indicate which results are theirs. If no such indication is made, results have been originally produced by the author.

³²<https://orcid.org/0000-0001-5123-4956>

Experiment 1: spatial attention and predictability

Short methods and results

Start: analyzed by Oussama Abdoun (end on page 136)

In a first step of the behavioral analysis, multiple sanity checks with respect to respective trial count distributions of conditions, trigger codes etc. have been performed. Additionally, it has been verified that the data is not subject to the Simon effect [408]³³. On average, participants responded faster and more accurately to easy, valid and congruent trials (compared to difficult, invalid, in-congruent), which verifies basic assumptions about how the task should work. Congruency here, is defined as whether the target and distractor have similar orientation. Furthermore, accuracy and RT data has been investigated with respect to attention condition (V100, V90s, V90b) and predictability (P5, P10, P80). It has been found that RTs for V100 are lowest, followed by V90b and V90s. Accuracy values for valid trials do not differ between any attention condition, similar to difficulty. However, invalid trials do expose a significantly lower accuracy for V90b as compared to V90s. Counter-intuitively, RTs for invalid congruent trials in V90b are unexpectedly low (see Figure 29 bottom left) as well as accuracy for in-congruent invalid V90b trials which drops to 53% (not shown in the figure). For predictability only a minor effect could be observed for accuracy and RTs of easy trials (see Figure 29 right).

³³The Simon effect [408] describes the tendency of participants to respond differently (in RT and accuracy) to trials where stimulus and response side are congruent as compared to in-congruent.

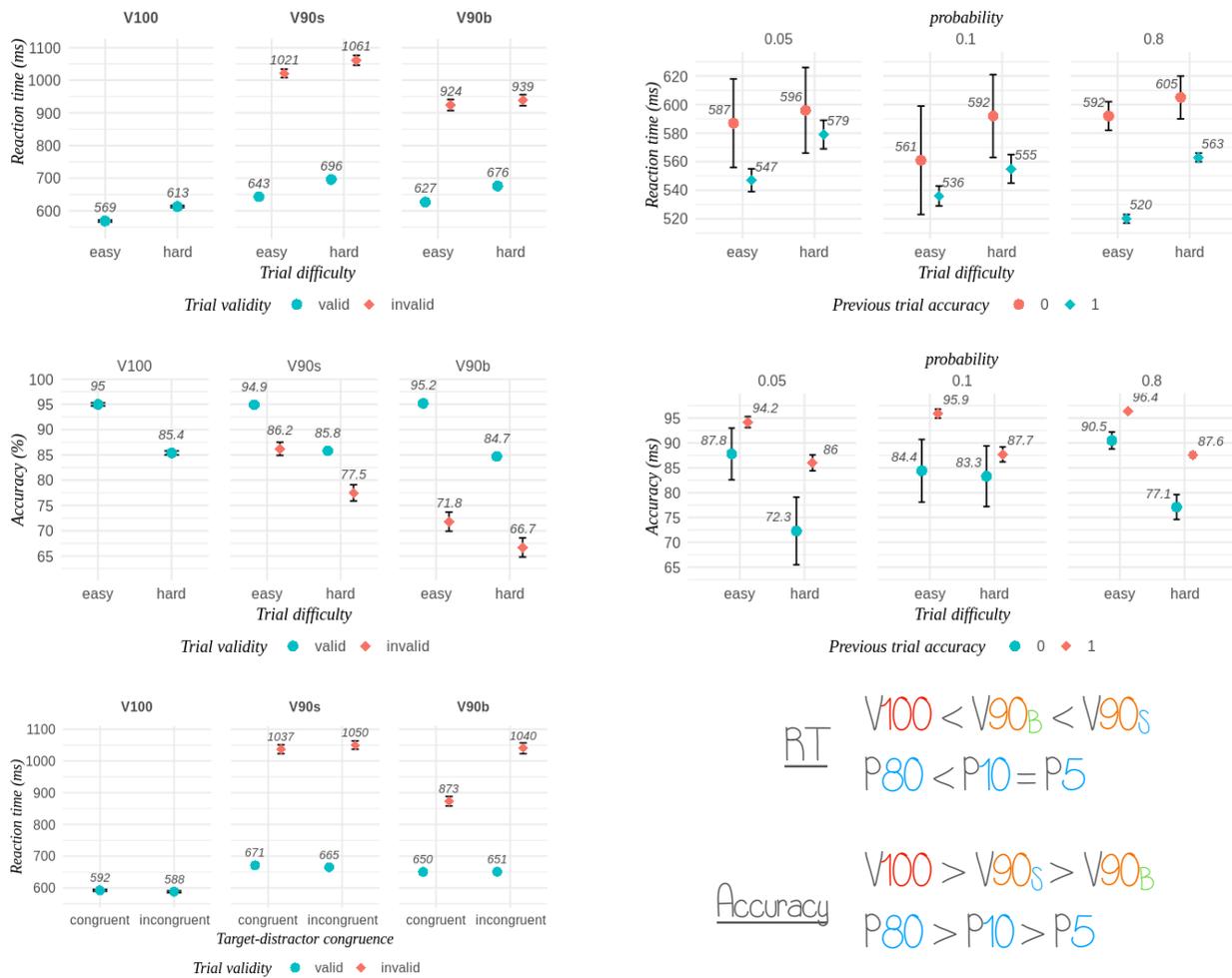


Figure 29: **Accuracy** and **reaction time (RT)** results for Experiment 1 (see page 123). Left: results with respect to the respective attention conditions V100, V90s and V90b. Right: results with respect to predictability. Bottom-right: summary. Reaction times for V100 were shortest, followed by V90b and V90s, for easy and difficult trials. In general, easy and difficult trials were different in reaction time (RT) and accuracy, where easy trials resulted in higher accuracy and shorter RTs compared to difficult trials. If target and distractor stimuli were congruent (same orientation), RTs for congruent trials in V90b were greatly sped up. However, accuracy for in-congruent trials in V90b dropped to 53% (not shown), indicating the relatively higher RT in the same condition might be due to delayed guessing. Accuracy of the congruent trial has not been affected, indicating that participants in V90b tended towards response repetition (not shown either). The effect on accuracy as a function of attention condition was larger in the in-congruent condition for hard trials for V90s followed by V90b. In general a tendency towards higher accuracy values for P5 followed by P10 and P80 could be observed although with minimal effect size ($\approx 1.5\%$) accompanied by a drop in respective RTs, at least in easy trials where the previous trial was answered correctly. This figure has been reproduced with kind permission of Oussama Abdoun.

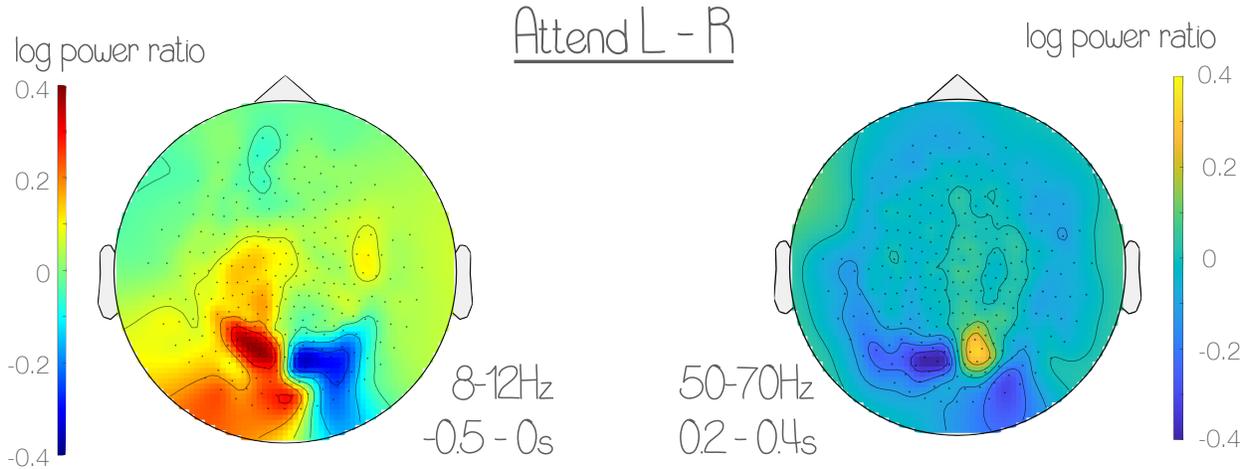


Figure 30: **Sensor level time-frequency data** for anticipatory α and stimulus induced γ band activity. Left: for the contrast between the attend left or attend right conditions an α band decrease (8 Hz to 12 Hz) has been observed for the pre-stimulus period (500 ms prior to stimulus onset until stimulus onset) over contra-lateral sensor sites (relative to the side of attention), whereas an increase is observed for ipsi-lateral sensors. The \log_{10} power ratio between baseline and respective time range of interest indicates ranges between $\approx \pm 0.4$ indicating the respective activity has been more than doubled (halved). Right: stimulus induced γ band response (≈ 50 Hz to ≈ 70 Hz) for a period of 200 ms to 400 ms after stimulus onset. A clear peak in the γ band between 200 ms and 400 ms after stimulus onset could be observed over contra-lateral occipital sensor sites as well as a corresponding ipsi-lateral negative deflection for the contrast between attend left and attend right. This figure has been reproduced with kind permission of Maryam Mostafalu [1].

In general, the MEG pre-processing pipeline follows what has been implemented by Ferez [407] (see Experiment 2 on page 143). Trials have been excluded if SQUID jumps or muscle artifacts had been detected by the automatic artifact detection algorithm, implemented in FieldTrip [324]³⁴. Only sensor level time-frequency analyses have been performed. For that reason, planar gradients have been computed [409] applying the nearest neighbor method. No statistical analyses have been conducted. Figure 30 depicts the result of a first sensor level analysis for α in the pre-stimulus period (500 ms prior to stimulus onset until stimulus onset) and γ for the early post-stimulus time interval (200 ms to 400 ms after stimulus onset). Both have been computed as the contrast between attend left and attend right condition. As expected, a clear peak in the γ band between 200 ms and 400 ms after stimulus

³⁴https://www.fieldtriptoolbox.org/tutorial/automatic_artifact_rejection/

onset could be observed over contra-lateral occipital sensor sites as well as a corresponding ipsi-lateral negative deflection. The respective increase peaks on a \log_{10} power ration scale between time period of interest and baseline at ≈ 0.4 which indicates an increase of more than twice the baseline activity ($10^{0.4} \approx 2.5$). Furthermore, an attention side specific α band response can be observed over occipital and parietal sensors sites in the pre-stimulus period as well. Thereby, the contrast between attend left and attend right condition is expressed as α band decrease over contra-lateral and an α band increase over ipsi-lateral occipital sensor sites. Both peak on a \log_{10} power ration scale at ≈ 0.4 , again indicating an in-/decrease of activity of a factor of more than two ($10^{0.4} \approx 2.5$).

End: analyzed by Maryam Mostafalu [1] (start on page 136)

Short discussion with respect to the present work Most importantly, the presented results have not been tested statistically and hence the respective discussion remains highly speculative. Nevertheless, based on experience, a \log_{10} power ratio value of 0.4 in a spatially limited region often indicates statistical significance in a corresponding test. Such test could e.g. be the cluster based permutation test for sensors level time-frequency data [352]. Without such a test, no final judgements can be made. Nonetheless, assumed to be "real" results, those findings would be in line with the literature. First of all lateralized attention α changes in occipital cortex areas have been observed in many previous studies for post-stimulus [251, 410, 411] and anticipatory α activity [412]. Thereby, occipital sensor sites contra-lateral to the attended stimulus expose a relative decrease, whereas ipsi-lateral sensors expose a relative increase [413]. The similarity between the literature and the presented results, combined with the relatively large power changes (\log_{10} ratio) is expected to finally culminate in a significant statistical test. For γ power changes, this is similar. Post-stimulus early trial γ power increases over contra-lateral and decreases over ipsi-lateral occipital sensors after stimulus onset (especially for gratings [189]) are in line with what has been expected based on previous literature [185, 238, 240, 414]. Furthermore, the relatively strong γ power de-/increase (\log_{10} ratio of 0.4) - limited to a small number of sensors - indicates effect validity. Since γ band oscillations have been related to feature specific processing [184, 240, 415], the respective orientation determination of the attended gratings might be what is reflected here. This would be in line with findings from the EEG-fMRI experiment that is part of the presented work (see "Feature specific neuronal oscillations in cortical layers (in prep)"), where a significant increase in γ band oscillations for early trials in a similar frequency range with similar stimulus material could be observed for the contrast between voxel that prefer a respective stimulus orientation over those that do not. This would hence indicate a potential relationship between γ band power and / or coherence changes [240] and / or reaction time (RT) and / or response accuracy (γ reflects ongoing stimulus processing which would influence behavior). Furthermore, the relationship between pre-stimulus α

band activity and RTs (and response accuracy), similar to what has been done for Experiment 2 on page 143 [407] could be performed. For Experiment 2, pre-stimulus α band power was related to RTs in one of the experimental conditions. It has been found that increased pre-stimulus α band activity over regions processing the respective distractor was larger for fast as compared to slow trials. Furthermore, β band activity in the pre-stimulus period exposed a similar effect in the ventral attention network (VAN), which has been linked to successful inhibition of distracting information at known locations [254, 402]. Here, it would be expected that a respective α activity change can be observed for the **V100** as compared to **V90** conditions as well as between **V90b** and **V90s** (see Experiment 1 on page 123). It has been found that upper α band power changes are related to whether a target stimulus or non-target stimulus was presented [416] and "involuntary" attention [417]. Since α power changes have also been related to stimulus specific pre-stimulus processes that have been hypothesized to reflect feedback directed activity and targeted distractor response modulation [91], a differential response pattern for the three visual conditions is expected. This is because during the V100 condition, no switch of attention side will happen (guaranteed) and hence full lateralized attention suppression of unattended stimuli would be reflected in contra-lateral α power decrease and ipsi-lateral α power increase [251]. During the V90 condition however, the subject is to some extent required to "monitor" the respective unattended side as well, because with a probability of 10%, the respective cued location will be invalid. The difference between the V90s ("switch") and V90b ("bottom") condition is, how the respective shift of attention is communicated. In V90s, the black squares indicating the target shift would *switch* location on the distractor *and* former target stimulus. Thus, the actual indication that the target is located on the other side is reflected on the former target stimulus. No attention monitoring of the unattended side is required during this condition with respect to the switch indicator. However, it is hypothesized that in order to improve performance in case of such an event, a small fraction of spatial attention is allocated to the unattended side. Even more attention is expected to be allocated to the unattended side in the V90b condition. Since both black squares indicating the target shift move to the *bottom* (no visible change of stimulus on the former attended stimulus), a shift of attention to the respective other side would only be possible if the unattended side is monitored for shift indication as well. Correspondingly, the respective α lateralization (see e.g. lateralization index [173]) should decrease, such that $\bar{k}_{lat}^{V100} \geq \bar{k}_{lat}^{V90s} > \bar{k}_{lat}^{V90b}$. In addition to that, respective variations in stimulus predictability need to be investigated as well.

Remaining work The remaining analyses for this experiment comprise multiple source analysis approaches, as well as connectivity analyses with respect to attention and predictability conditions (see Experiment 1 on page 123). A first step however would be the in depth analysis of RTs and accuracy and relating those to the MEG data. RTs and accuracy are expected to be highly dependent on the condition. Preliminary RT analyses performed

by Oussama Abdoun (and initially Lucie Rissoan, an intern under the author's supervision), revealed significantly lower RTs for the V100 compared to the V90 condition as well as a significant difference between valid and invalid cue for both V90 conditions combined (both $p < 0.05$). This was expected, since no monitoring of the unattended side whatsoever has been required in the V100 condition which potentially enabled participants to focus on stimulus processing alone. The significant difference in RT between valid and invalid cue conditions in V90 can be explained by the actual switch that has to be performed for invalid trials which in itself requires time to be executed, irrespective of the underlying features. In addition to that, RTs have been shorter for V90b than V90s. In fact, the reverse pattern has been expected, since more monitoring time capacity is required in V90b. On the other hand, since the accuracy for invalid trials is much lower in V90b, the longer RT in V90s seems to pay off (see Figure 29). In addition to that, during V90s the only source of information is the attended side which potentially causes a locking mechanism of attention to that side, making it harder to switch [179] (consequently longer RTs). Further analyses revealed that accuracy for invalid trials in V90b that required a change of response compared to what has been anticipated based on the former target stimulus drops significantly (53%). In addition to that, RTs for congruent trials in V90b are suspiciously sped up compared to in-congruent trials. Taken together, those findings indicate that participants either tend towards guessing for invalid trials in V90b especially in the hard condition or miss the switch indicator in almost half of the invalid trials. The influence of the target is presumably relatively large, because congruent invalid (same response button) trials receive higher accuracy and lower RT values compared to in-congruent trials. Compared to V90s, accuracy for V90b is significantly lower for invalid trials and again taken together with the shorter response time in V90b indicates a higher rate of guessing (shorter RT, lower accuracy) or rate of missing (and then responding to the "false" target).

Effects of predictability on accuracy are relatively small ($\approx \pm 1.5 \%pt$) with highest predictable stimuli receiving highest and lowest predictable stimuli receiving lowest accuracy values. RTs thereby do not differ between prediction groups, however the variance in the 5% and 10% predictability groups is much larger for trials that are preceded by a trial that was answered incorrectly as compared to the 80% predicted trials. However, since 80% predicted trials naturally occur $8\times$ as often as the other two groups (10% and $2 \times 5\%$), their respective trial count is much lower and hence the variability might be greater. For easy trials, a tendency for a negative correlation between predictability and RTs is observed (see Figure 29 top right). However, results for hard trials with respect to RTs remain slightly more puzzling (especially when taking into account previous response correctness see Figure 29) as the most predictable exposed very high RTs compared to the general trend. I hypothesize that those prolonged RTs are perhaps caused by participants doubting their perception. Since hard trials have been specifically designed to be more ambiguous with respect to the orientation than easy trials and given that for each set of predictability per stimulus one

combination always points back to itself (predicts itself with 80%), participants potentially hesitate in this situation. The ambiguity and the possibly implicit assumption that so many repetitions of the same orientation would be unlikely might inflict hesitation and double checking, which would increase RT. So far this has not been specifically tested. Whether participants could explicitly predict a respective stimulus has however been assessed at the very end of all recording sessions. Each of the four stimulus categories was paired with each of the other four stimulus categories and participants were explicitly asked to predict which stimulus would follow the current reference stimulus. This data however has not yet been analyzed. Irrespective of that, it could be demonstrated that participants are able to implicitly - or unconsciously - learn statistical relationships without being able to formulate those relationships explicitly [418]. Similar to the implicit association task (IAT) [419]³⁵, where implicit associations lead to faster RTs and is reflected in an increase of RTs if behavior is modified with respect to underlying initial causes, faster responses are expected for stimuli that could be predicted with highest probability (i.e. 80% in the present case). Thereby, a violation of prediction (20% of cases) would lead to slower RTs, because it would require correction. In the present experiment a condition was added, where stimulus expectation is violated but the resulting button response the same (10% cases). Additionally, 2 × 5% of trials not only violated the respective predication, but furthermore also the predicted button response (combined 10%). Note, that theoretically the prediction with respect to difficulty could still be valid for one of the two final cases, however it is hypothesized that the violation in predicted orientation overrules the respective decision whether the difficulty is similar or different. Those three major conditions (predicted with 80%, predicted with 10% or predicted with 5%) are expected to consecutively cause an increase in RT. Here, only the attention effect shows a clear separation between the experimental groups. A tendency for a predictability effect (especially for easy, predicted trials) has been observed as well, however less clear cut as expected. One hypothesis would be the relatively high task complexity that explains this. High task difficulty would make it more demanding to a) detect that there are indeed four different stimuli and b) that those follow a statistic order. If each trial is so difficult that all resources are required to perform the task, statistical learning might be hampered. An in depth analysis has to be conducted to see if the predictability effect is shadowed by e.g. task difficulty (individually determined).

Beside in depth analyses on behavioral responses (that can be later related to MEG data), neuronal responses in form of MEG data need to be fully investigated on various levels. In a first step, pre- *and* post-stimulus α and γ oscillations have to be contrasted for a variety of conditions. First of all the attend *left* compared attend *right* contrast needs

³⁵The implicit association task (IAT) measures implicit associations between categories by differences in RTs [419]. It is hypothesized that "true" associations cause behavior that is carried out more or less heuristically / automatically and if top-down control is exerted (e.g. to hide some implicit associations), this is reflected in increased RTs.

to be transformed to reflect "attended side" and "not attended side", by combining both. The respective contrast furthermore needs to be performed for time-frequency transformed sensor and source level data, similar to what has been done in Experiment 2 (see page 143). It is expected that preliminary findings can be confirmed for the pre-stimulus period in α and that those can be observed during the post-stimulus period as well [251]. For both, pre- and post-stimulus period, event related α power decreases are expected for contra- and α power increases for ipsi-lateral regions. In the γ band the respective post-stimulus γ band increase is expected to be confirmed statistically and will locate to visual cortex areas (V1-4) for this kind of stimuli (see e.g. "Feature specific neuronal oscillations in cortical layers (in prep)"). However, a pRF mapping [72] performed using the respective fMRI data, could help to determine respective primary visual regions for a ROI analysis. Furthermore, the functional localizer that has been recorded at the end of session 4 (see Figure 26) could help to further narrow down ROI analyses to particular areas. For both, α and γ , it is expected that findings from the EEG-fMRI experiment can be replicated (low α and high γ power in V1). Effects in the γ band are expected to be highly attention and hemisphere specific, with higher responses for attended stimuli in visual cortex regions contra-lateral to the attended stimulus [415], which has been confirmed here. Furthermore, this would explain the positive correlation between the γ band signal and generally deactivated voxel (possibly not attended) found in the EEG-fMRI experiment on trend level.

In addition to that it could be demonstrated that FEF exerts influence during attention on V4 and V1 [76, 244, 245, 254] which could be linked to changes in α band connectivity [251]. The influence of higher order frontal and PFC regions on primary visual regions [78] has further been reported to be modulated in the θ band [204, 420]. In Experiment 2 (see page 143), pre-stimulus θ band increases were found to be related to behavioral performance. In-congruent trials in one of the two attention conditions, exposed a stronger θ band (3 *Hz* to 7 *Hz*) increase in activity over frontal areas, for fast compared to slow RT trials (median split). This indicates that changes in frontal θ band power during the pre-stimulus period, is related to fast stimulus encoding under conditions where the target information is rivaled. However, it remains to be investigated whether those θ band power changes translate to changes in top-down connectivity as well, as they could furthermore be explained by θ power changes with respect to the decision level of cognitive control [421]. Analyses with respect to cortico-cortical connectivity changes could be carried out using a DCM model [155, 157] or more traditional approaches, such as mutual information or Granger causality [152].

While still under development, which is led by James Bonaiuto, a new laminar level MEG processing pipeline allows for the investigation of neuronal oscillations on the level of cortical laminae using MEG. It could be demonstrated that sensori-motor related β bursts can be mapped to two cortical layers [220] as well as low and high frequency activity in the

visual cortex [268]. This method allows for the opportunity to directly investigate neuronal oscillations on the level of cortical laminae, non-invasive in humans and test predictions from very low scale models, such as the canonical microcircuit model by Bastos et al. (2012) [82]. Provided the presented task (see Experiment 1 on page 123), attention related changes to cortical α activity are hypothesized to be reflected mostly by deep layer activity [52], whereas changes in γ band power is hypothesized to locate to superficial layers [52, 238] (but probably deep layers as well [80, 272]) of primary visual regions. Thereby, a communication channel between processing regions is thought to be reflected by an α decrease that is shared coherently over jointly active regions. Depending on the respective side of attention or cue validity condition (V100, V90s, V90b), those changes are expected hemisphere specifically in primary visual areas due to the lateralized presentation [251], and in higher order regions due to target PFC or FEF activity [78]. Specifically, it is expected that attention related α band hemispheric lateralization and fronto-occipital connectivity changes as a function of attention condition. It is expected that the lateralization and connectivity between primary visual areas and frontal areas is strongest for V100, followed by V90s and V90b. Respective power connectivity changes are again expected to locate to deep layers. Moreover, a potential influence of frontal regions on primary visual areas is to be expected in the θ band, exposing a similar response pattern as described for α [204, 420]. The canonical microcircuit model (see "Canonical microcircuits: cortical compute modules") further suggests a reflection of prediction errors in superficial layer γ [82, 246]. Thereby, a violation of prediction should be reflected in stronger γ band power than for valid predictions [422] which is expected to locate either to superficial layers [52] or superficial and deep layers [80, 272]. Violations thereby can happen on two levels. First of all the cue could be valid (80%), invalid but with the same response button (10%) or invalid but with a different response button required ($2 \times 5\%$). Hence, γ band power in superficial layers should be larger for trials violating certain predictions and even larger (or across more widespread network activity) when the prediction violation requires a change in behavior compared to what has been predicted. Thereby, increased γ band activity is expected to be coherent between "active" regions (e.g. V1 and V4) [52, 240]. In turn, β band power is thought to reflect the actual "predictions" that are made in top down direction conveyed by coherent α [219]. Possibly those can be extracted in the sense that respective trials could be decoded based on the *expected* stimulus, which will be investigated by Maryam Mostafalu [1] as part of her PhD project. Similar to respective α and γ band analyses, θ and β band responses have to be investigated with respect to attention selectivity, predictability, cortical layer specificity and connectivity.

Experiment 2: feature attention and conflicting information

Start: analyzed by Maxime Ferez [407] (end on page 146)

Short methods In a first step, trials were excluded if the respective response time was shorter than 300 *ms* or longer than the average and 2 standard deviations ($\mu + 2\sigma$). Remaining trials have been transformed into epochs for a time interval between -700 *ms* and response time or at the maximum 1 *s*, relative to the stimulus onset. An automatic artifact rejection algorithm based on FieldTrip [324] ³⁶ was applied to reject trials containing SQUID jumps and muscle artifacts. Furthermore, trials have been rejected based on the eye tracking data. Trials containing saccades or blinks or trials where the subject exposed weak central fixation performance have been furthermore rejected. Using this procedure, 23.6% of all trials have been rejected. One subject was excluded due to the overall density of artifacts in the signal. For the sensor level analysis, planar gradients have been calculated [409] using the nearest neighbor method. Before the sensor level time-frequency analysis has been conducted, the data was band-pass filtered for frequencies between 3 *Hz* and 40 *Hz*. The calculation of the power spectrum has been conducted using individual frequencies for each Hanning taper, such that exactly three cycles fit the sliding window length. Source level time-frequency statistics have been computed using a dynamic imaging of coherent sources (DICS) beamformer [423] or LCMV beamformer, where the later is transformed into time-frequency data on the level of reconstructed virtual channel signals. A single shell model derived from anatomical MRI data has been used to compute the forward model [424]. Thereby, the brain space has been parcelized into voxel (1 *cm* isometric) and morphed towards the standard MNI brain [425]. All respective source level data has been contrasted between conditions for frequencies in the α range.

A repeated measures analysis of variance (ANOVA) with factors condition (attend-face or attend-name), congruency (congruent or in-congruent) and type of stimulus (each of the face/name combinations separately) has been conducted for a reaction time (RT) and accuracy analysis. Violations of sphericity have been corrected using the Greenhouse-Geisser [426] correction. Post hoc testing was done using the Tukey-Kramer test ($\alpha = 0.05$). Both, sensor and source level time-frequency analyses have been statistically tested using a cluster permutation test [352] using 1000 permutations to build the permutation distribution.

³⁶https://www.fieldtriptoolbox.org/tutorial/automatic_artifact_rejection/

Short results Analysis of RTs revealed a difference between the attend-face and attend-name condition ($F(1, 39) = 354.07, p < 0.001$), where average (σ) reaction times for attend-face were 629.1 ms (138.6 ms) and for attend-name 565.2 ms (157.4 ms). Additionally, RTs were shorter for congruent as compared to in-congruent trials ($F(1, 39) = 161.9, p < 0.001$). A significant interaction between congruency and condition was observed ($F(1, 39) = 68.0, p < 0.001$). Post hoc analyses revealed that the difference between congruent and in-congruent trials with respect to RTs was larger in the attend-face condition than in the attend-name condition ($p = 1.06e - 6$ and $p = 4.28e - 7$ respectively). See Figure 31 for a graphical representation of the RT results. Accuracy analyses revealed a significant difference ($F(1, 39) = 45.1, p < 0.001$) for the average accuracy between attend-face and attend-name conditions (94.7% and 91.5% respectively). An additional main effect was observed for congruency ($F(1, 36) = 145.3, p < 0.001$), with in-congruent trials exposing lower overall accuracy.

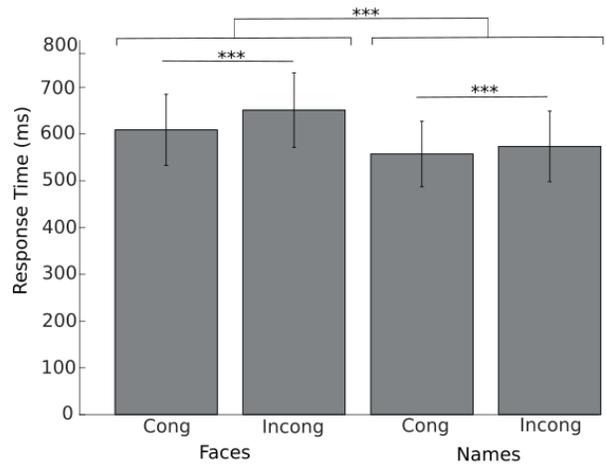


Figure 31: **Reaction time per condition** On average reaction times are significantly smaller for congruent (Cong) than in-congruent (Incong) stimuli as well as trials in the attend-names (Names) than attend-face (Faces) condition (***) indicates $p < 0.001$). This figure has been reproduced with kind permission of Maxime Ferez [407].

An anticipatory α effect has been observed during the pre-stimulus period (400 ms to 50 ms prior to the stimulus) for a cluster spanning fronto-temporal sensor sites for the attend-face compared to the attend-name condition ($t_{sum} = 729.84, p = 0.041$) for frequencies between 8 Hz and 14 Hz (see Figure 32 left). Source level analysis revealed a significant cluster for the same frequency band ($t_{sum} = 1.25e3, p = 0.008$) over left occipito-parietal and temporal regions (see Figure 32 right). More precisely, this includes left occipito-temporal (LOT) (possibly including posterior VWFA), left medio-frontal (LMF), left pre-motor (possibly motor preparation in the word condition because RTs in that condition are significantly lower in general) and bilateral parietal areas.

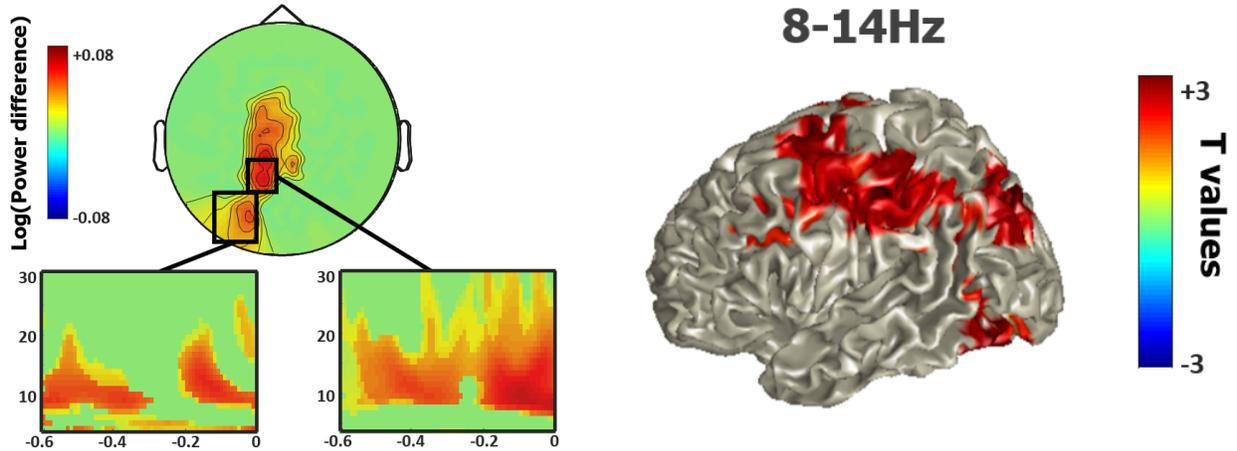


Figure 32: **Increased α band baseline activity** in the attend-face compared to attend-name condition. Left: A significant cluster over left occipito-parietal sensor sites for the comparison between attend-face over attend-name condition has been observed for frequencies between 8 Hz and 14 Hz . Time frequency plots are obtained from the average activity in encircled sensors. Only significant clusters ($p < 0.05$) are shown. Right: Source level analysis for frequencies between 8 Hz and 14 Hz . Red areas mark a significant α power increase for attend-face compared to attend-name ($p < 0.05$). This figure has been reproduced with kind permission of Maxime Ferez [407].

Additionally, α band power changes have been related to performance. If α band increases - as hypothesized - reflect functional inhibition of neuronal populations [251] tuned to process conflicting stimulus information [91], then α band power changes should be related to RT (better inhibition of distracting information or less sampling of those would lead to lower RTs). This is precisely what was found. In a frequency range between 8 Hz and 14 Hz for a trial period of -400 ms to -50 ms relative to stimulus onset, an increase for fast over slow trials (split at the median RT) in the attend-face condition was observed for incongruent trials ($t_{sum} = 1.12e3, p = 0.009$). See Figure 33 left. No difference has been found for the attend-name condition. Source level statistical analysis revealed a significant cluster for posterior VWFA ($t_{sum} = 317.74, p = 0.009$). In addition to that, a significant θ band increase over frontal areas ($t_{usm} = 3174, p = 0.01$) has been interpreted as sign of cognitive control [427] and here faster RTs.

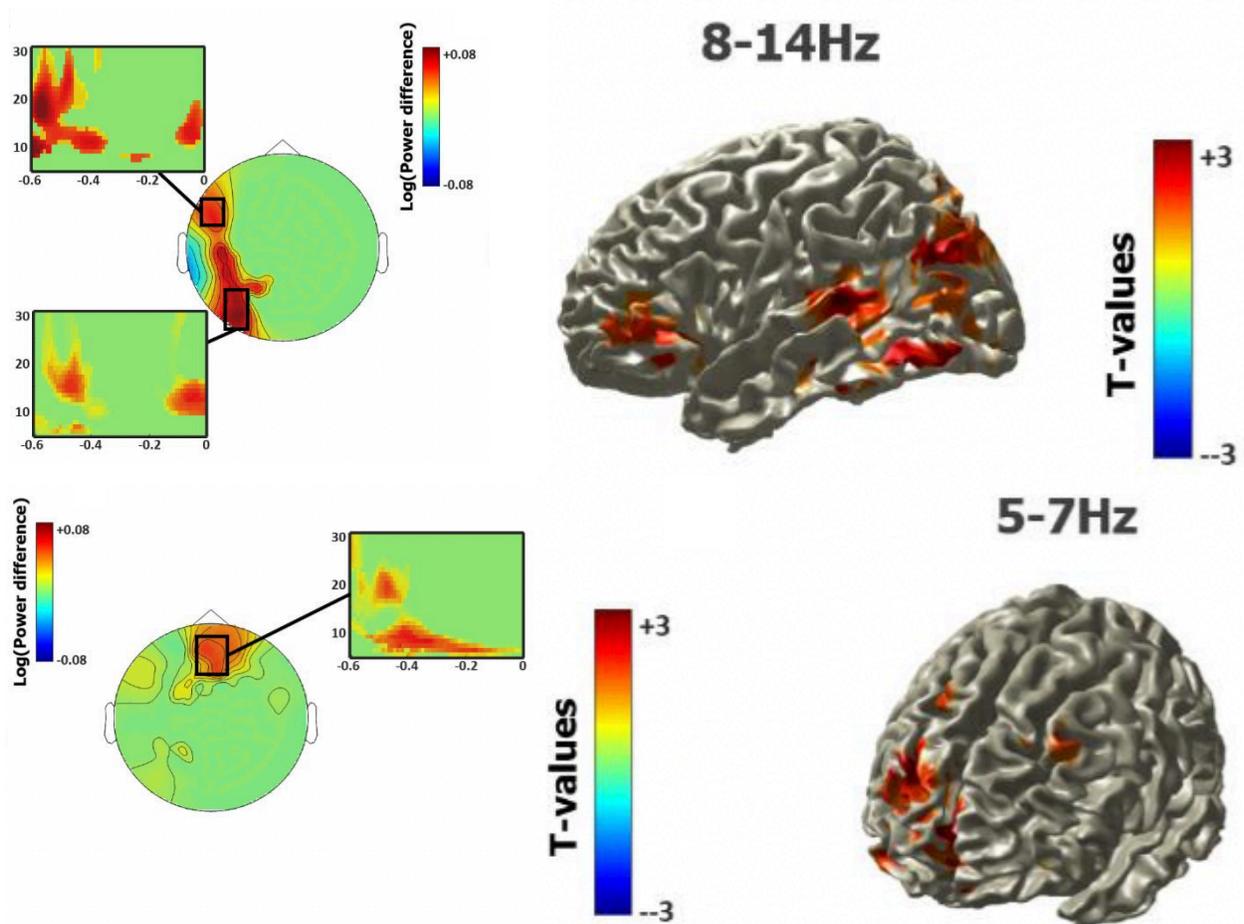


Figure 33: **Increased α and θ band activity for fast over slow RT trials** (median split) has been observed in a range between 8 Hz and 14 Hz (top) and 5 Hz to 7 Hz (bottom) respectively during the pre-stimulus period for in-congruent trials within the attend-face condition. Left: Significant cluster ($p < 0.05$) over parietal, left occipito-temporal and frontal sensors for frequencies in the α and over frontal sensors in the θ band. Time frequency plots are obtained from the average activity in encircled sensors. Only significant clusters ($p < 0.05$) are shown. Right: Corresponding source level data for the period from -100 ms prior stimulus onset until stimulus onset. Red areas indicate significance ($p < 0.05$). This figure has been reproduced with kind permission of Maxime Ferez [407].

End: analyzed by Maxime Ferez [407] (start on page 143)

Short discussion with respect to the present work Experiment 2 has been conducted in order to investigate stimulus type dependent α band oscillations (see e.g. "Research question"). It has been hypothesized that feedback related α band oscillations coherently decrease in power over regions that are used for joint stimulus processing. This however would be an anticipatory process in order to recruit cortical areas relevant to stimulus processing [91]. Above results [407], combined with further results from the corresponding paper in preparation [428] partly confirm this hypothesis. During the pre-stimulus period (400 *ms* to 50 *ms* prior to stimulus onset) a significant increase in α band activity has been observed for the attend-face over the attend-name condition for parts of the VWFA. This is in line with the hypothesis that α band activity increases over stimulus processing specific areas that interfere with target processes [233]. The presented data exposes such a targeted α band increase over VWFA, where the *word* component of the stimulus would be a distractor that needs to be suppressed. Not only would this be in line with the GBI, but furthermore predictions derived from the work of Bonnefond et al. (2017) [91]. Similar to spatial attentional sampling [180], where higher α amplitudes are interpreted as modulator causing shorter "window of opportunity" periods for information to be sampled, a targeted α band increase over VWFA during facial attention would potentially limit the distraction caused by in-congruent stimulus features. Indeed, RTs are longer for in-congruent as compared to congruent stimuli. While this finding is fundamentally trivial and results as a consequence of conflicting information, similar to the original Stroop task [429, 430], the simultaneously occurring α band increase indicates the modulation of irrelevant information processing.

In addition to that, all in-congruent trials from the attend-face condition have been split according to the median of the RTs into *fast* and *slow* trials. Fast and slow trials have been compared, revealing a significantly increased α band activity during the pre-stimulus period for fast as compared to slow trials. This indicates that high α power is related to a faster ("better") stimulus processing due to successful inhibition of potentially confounding information from the VWFA in the attend-face condition. Nevertheless, no particular α power *decrease* over VWFA for the attend-name condition has been observed, which would have been predicted as well. However, since the lexical information seems to interfere more with attention in the attend-face condition, as compared to the reverse (facial information less distracting in attend-name condition), the effect could just be less prominent. Recent findings explicitly link the amount of distraction that is delivered by interfering information to the strength of α power increases [233]. Furthermore, the observed activity for both contrasts (attend-face compared to attend-name and fast compared to slow trials) has been widespread over parietal areas. In general it has been questioned, whether α band activity indeed reflects cortical inhibition [176]. Recently it has instead be suggested that parietal α reflects a gating mechanism, modulating the flow of information [175, 431]. In addition to that it has been reported that anticipatory α power changes only occur during the anticipation of stimuli that can be predicted in terms of spatial location [432]. Here, this would be the case, since word

length, position, font size and capitalization had been standardized. Perhaps information is indeed gated for predicted distractor locations in lower level visual areas, at least within the attend-face condition.

Remaining work Most importantly, condition specific single trial responses need to be analyzed for the stimulus presentation period (very preliminary results obtained by Mathilde Bonnefond - which are not reported here - tend to confirm this). Both, sensor and source level analyses as presented [407, 428] need to be extended to the main stimulus processing period (e.g. 0.1 s to RT after stimulus onset). It is expected that - irrespective of the condition - a post-stimulus α band decrease over occipital areas and sensor sites can be observed, as well as a γ band increase over the same areas. Both are expected to be functionally related to stimulus processing, where α has been found to modulate bottom up information [264] and γ to reflect ongoing neuronal activity in active cortical regions [238]. However, in order to investigate primary visual regions of interest (e.g. V1-4), retinotopic field mapping fMRI data has to be analysed. The procedure thereby is similar to what has been done in the laminar level EEG-fMRI experiment (see "Feature specific neuronal oscillations in cortical layers (in prep)"). A GLM is used to related spatial resolved stimulus patterns on the screen to voxel activity from the fMRI [72]. This procedure allows the functional identification of (at least) primary visual areas. Furthermore, a localizer fMRI experiment, targeting FFA using facial stimuli, has been conducted. Respective primary and higher order visual areas could be targeted in a region of interest (ROI) analysis to investigate respective α (and β) and γ band changes. Furthermore, the MEG localizer tasks for face and name (see page Experiment 2 on page 121) could be furthermore included to narrow down functional ROIs. Beside primary visual area α decrease and γ increases, a similar pattern would be predicted for the respective targeted area (e.g. FFA or VWFA). This hypothesis is based on previous findings on γ , where attended facial stimuli elicited higher γ band responses in FFA; and α , where FFA related activity was associated with lower α power [433].

Furthermore, functional and effective connectivity could be estimated via a DCM model [155, 157] or more traditional approaches, such as mutual information or Granger causality [152]. Based on several frameworks, α band connectivity should be directed in feedback direction from higher order visual regions to lower order visual regions, whereas γ oscillations are thought to be associated with feed-forward directed activity [52, 82, 91]. If α band oscillations indeed setup communication channels in feedback direction via inter-area coherence, then not only should α band power decrease over areas relevant to stimulus processing with feedback directed information flow, but furthermore α band activity is expected to be phase coherent between high and low order areas [363]. Ideally, a respective feature attention dependent increase or decrease over VWFA or FFA in α power respectively could be observed for the attend-face condition and the reverse for the attend-name condition. Again, for this

purpose the pRF mapping and MEG localizer data needs to be analyzed in order to obtain structural and functional ROIs.

In addition to that, the laminar level MEG processing pipeline is currently under development, led by James Bonaiuto. It could be demonstrated previously that sensori-motor related β bursts can be mapped to two cortical layers [220]. In addition to that it has been demonstrated that laminar level MEG is in principle possible for visually evoked activity too [268]. The resulting two layer model hence allows to repeat previously mentioned ROI specific source space analyses at the level of cortical laminae. It is expected that feedback directed α band activity can be found predominantly in deep layers, whereas γ band activity has been associated mainly with superficial layer activity [52, 91]. However, since forward directed pathways have been located in layer 5 as well [80], γ band activity might be related to both layers. On the most in depth level, it would be expected that higher order primary regions (e.g. V4), modulate lower order primary region activity (e.g. V1) in the α band via deep layer connections [52]. However, an additional attention related α component (either spatial or feature specific, but probably both) would further be expected to gate (ir-) relevant information [309, 431]. The attention related α component would hence gate respective task specific (ir-) relevant information on a higher order level (e.g. (dis-)inhibiting FFA or VWFA respectively) and a more stimulus processing related α would set up necessary communication channels via coherence. Specifically, high α coherence between stimulus processing early visual regions (V1-4) and higher order regions (FFA in the attend-face [434] and VWFA in the attend-name condition [435]) is expected. In addition to that high γ band coherence between low and high order visual regions for areas that are targeted (e.g. FFA in the attend-face condition) as compared to those that are not (e.g. VWFA in the attend-face condition) is expected to increase [240], but mainly due to its nesting in coherent α connectivity between e.g. V1 and FFA [91, 379]. In this case α band activity over VWFA possibly exposes an anti-phase relationship with FFA α and increased power.

Lastly the investigation needs to be extended to the θ and β frequency bands. A first intuition about expected β band results can be derived from the pre-stimulus data that has been presented above [407] and the respective publication in preparation [428]. Over areas corresponding to the ventral attention network (VAN), a significant increase in activity could be observed for the attend-face compared to the attend-name condition for a frequency range between 14 Hz and 30 Hz . Suppression of VAN activity has been related to the successful suppression of unattended (but relevant) stimuli and the resilience to attention being captured by unattended stimuli [254, 402]. The canonical microcircuits model by Bastos et al. (2012) [82] predicts that β band activity is related to top-down predictions. In the light of the given task, bottom up feature processing in general might be considered easy and respective distractor suppression more difficult. Here, pre-stimulus β activity might be explained in terms of which areas and / or features to suppress in order to successfully encode

the stimulus. However, this remains speculative since actual stimulus or feature predictability was not manipulated. See Experiment 1 on page 126 where stimulus predictability has been varied. With respect to θ , the above mentioned paper in preparation [428], that has been based on the here presented results [407] (see page 143), provides a first intuition. Ferez et al. (in prep [428]) found that increased θ band activity during the pre-stimulus period was related to the RTs. Trials that participants responded to with a fast RT thereby elicited significantly stronger θ band activity (5 *Hz* to 7 *Hz*) over frontal areas, including bilateral anterior cingulate cortex (ACC) and bilateral superior frontal cortices in in-congruent trials of the attend-face condition. This implies an involvement of frontal cortical control networks [436]. Potentially, θ plays a mediating role supporting the resolution of conflicting information. In addition to that α seems rather burst-like (see Figure 32 left, the left time-frequency plot to get an intuition). While not statistically verified yet, this burst like behavior has been suggested recently [232] and possibly relates to attention sampling [207]. Interestingly, here burst seem to occur every 200 *ms*, which would correspond to a 5 *Hz* rhythm. Whether a statistical relationship between 5 *Hz* θ and possibly phase gradient dependent nested α bursts [210, 211] exists, remains to be tested.

2.4 Artificial neural networks

Yet again, the global pandemic hindered most of the progress on this part of the project (see "COVID 19 remark"). For both major experiments (see below), previously recorded MEG has been a pre-requisite. However, it should not be a secret that this part of the project was treated as "bonus" right from the beginning. Only once the required MEG data has been collected sufficient training data would have been available to attempt potential ANN architectures seriously with respect to the research questions (see "Research question"). After MEG data analyses would have been finished, this part could finally be seriously addressed. This is due to the fact that for Experiment 1 (see page 154), MEG data is either *the* training data or required to compare the ANN to. Not only the raw data is required however, but moreover source level and time-frequency transformed data as well. For Experiment 2 (see page 155), MEG data is not inevitably required, however in order to implement the hypothesized cortical computational framework as accurately as possible, respective MEG results from both experiments are required. For those reasons, only very simple first steps with respect to potential network architectures have been undertaken.

Expected publication output

One first author publication has been planned on spatial features learned from spatially distributed γ band power by a CNN (Experiment 1, see page 154) and one first author publication on oscillatory activity and related "behavioral" changes in a trained SNN.

Introduction

Modern, highly sophisticated deep neural network (DNN) applications are based on simple models derived from findings about computational principles of the cortex. A first machine incorporating "cortical" compute principles was the *Perceptron* (Rosenblatt, 1957, [73]). It was based on the idea that each neuron in a biological network transmits neuronal signals (action potentials) to other neurons which has an effect on the later neuron, making it more or less likely to produce an action potential itself. The positive or negative impact a neuronal signal has on other neurons, depends - among many other cellular processes - on the respective neurotransmitters and the charge and amount of ions that are exchanged with the extracellular space. Hence, a simplified model of cellular communication could be expressed by the activation of a neuron that outputs a weighted signal (\pm) to one or more other neurons. The *Perceptron* was not a computer according to modern day standards, but rather a fully integrated machine that has a grid of input receptors that would output their signal onto a single output neuron. Given a certain threshold, the output could either be 0 or 1, enabling the Perceptron to separate between two classes. Thereby, each input to output

connection received a weight that was modified by element wise multiplying the input with the difference between predicted and target class (either -1 , 0 or 1) and adding the result to the already existing weights. This machine could - as a result - separate two classes of input examples (simple shapes) reliably, but failed at slightly more complex stimuli. It turned out that adding more and more - so called "hidden" - layers between input and output neurons (hence *deep* neural networks) greatly improved performance. Findings from Hubel & Wiesel (1962) [30] later inspired the theoretical development of the *Neocognitron* in 1982 [134]. Thereby, multiple layers of simple (or s cells) and complex (or c cells) are wired up such that in each processing stage, feature specific s cells respond preferably to the presence of their preferred feature and c cells if one or a specific combination of s cells are active. Thereby, features get grouped and are re"interpreted" as combination of simple and hence complex features. For example, basic features such as lines that are responded to by s cells are combined by c cells into corners. Higher order s cells in turn would respond to lower order c cells and so on. This idea later inspired what is known today as convolutional neural network (CNN) [135]. The brain has been a source of inspiration since day one of artificial intelligence (AI) research. However, only little effort has been undertaken to transfer insights from AI back *to* neuro-science.

As has been pointed out that a simple back-transfer of structural developments from artificial neural network (ANN) models that have been developed with a somewhat engineering perspective, such as error back-propagation [114], cannot transferred easily [86, 87]. Nonetheless, I would argue that potential difficulties in a respective back transfer of findings from AI to neuro-science are mostly limited to structural aspects. The respective back-propagation algorithm for instance could be shown to be *functionally* similar to what is theorized to be computed in dendritic connections, however based on a fundamental different "software" implementation [85]. The discovery of new ways to implement more powerful neural networks does not lead to any new insights into computational functions of the brain, however deriving hypotheses about underlying brain computations from ANNs where architectural similarities *do* occur, could potentially provide a powerful tool for a more in depth understanding of the brain from a computational perspective. Over the past ≈ 5 years, some researchers have focused on identifying exactly those similarities. The basic principle behind this kind of research is rather simple: it boils down to training one or more humans and a ANN on some (often visual) task and compare respective brain activity to ANN layer activations. Similarly, an ANN can be set up to roughly mimic the primary visual compute architecture and trained to produce pseudo-neuronal activation from original human brain data. This has been done e.g. by Seeliger et al. (2021) [124] in an impressive demonstration how powerful this method indeed is. Human subject fMRI data has been collected, while the subject was watching more than $23 h$ of a TV series inside the scanner. Afterwards, a CNN was built to mimic core parts of the visual processing pipeline: Retina, LGN, V1-3, MT and FFA. During training, the CNN received a temporally down-sampled version of the original

video stream as input data and was trained to replicate the respective fMRI data at each layer. Layer weights of the trained network have been extracted afterwards. Among rather cryptic patterns, many patterns that have been extracted from "V1" of the CNN (the first convolutional layer) exposed stripy, grating-like shapes. Applying the learned filter kernels to images resulted indeed in an edge enhanced version of that image. V1 is known to respond to different bar orientations and movement directions [30] and hence the assumption of a functional similarity between the brain and CNNs imposes itself. Nonetheless, this study does not *prove* that any of this is indeed what is computed, but that an approximate structure, trained to transform similar input patterns to similar output patterns learned features in lower layers to which lower hierarchy brain areas preferably respond. From a long list of potential issues of drawing the proposed conclusion from the presented results, possibly the most impacting are spatial and temporal resolution of both, the CNN and fMRI data. While the CNN does not capture any temporal dynamics, fMRI does on a very low temporal scale (order of seconds). Computationally relevant time scales for brain activity changes can very well exceed the $100\times$ per *s* order [146]. Seeliger et al. (2018) [126] related MEG activity to the activation profile of a CNN exposed to a similar task: object recognition. Again a CNN to brain convergence could be observed. Thereby, the highest correspondence between early CNN layers is to really primary visual region signals, early after stimulus onset, whereas the correspondence with higher order regions of the brain and deeper CNN layers is higher later in the trial. This demonstrates as well how - to a certain extent - ANN models can serve as a reference frame for human brain investigation. Again, thereby not so much structural, but more functional aspects seem somewhat comparable.

A full neuronal activation map has been recorded from the *Hydra* and yet the neuronal code that explained the *Hydra's* behavior could not be decrypted [24]. Deciphering the neuronal code however is not necessarily the only way how information about the *exact* computations of the brain could be derived. Again the potential methodology could be derived from AI research. While the exact underlying logic within a neural network is hidden to the observer, the collectively (from the neurons) computed functions *are* to some extent accessible. The aforementioned filter kernel visualization can count as such an example. The mathematical operation the network performs thereby is known (convolution), whereby the operator is learned and can be visualized in terms of feature preference by tracing back errors and generating a stimulus that maximizing the response at each layer [437]. This means that even though the metaphysical *why* the neural network arranged itself in a certain state remains hidden, *what* that state or function is, can be read out. Much has e.g. been hypothesized about the *exact* computational role of α oscillations [91, 224, 291], but only recently the relationship between brain data and functionally more or less similar artificial networks has been targeted. So far, even less investigated remains the relationship between ANNs and neuronal oscillations in terms of functional similarity. If neuronal oscillations indeed execute network function relevant signal modulations, then those respective functions

can potentially be either learned from the data and later visualized [124] or a trial by trial analysis of time-frequency transformed data reveals insight into *what* is computed, similar to what has been done using time-locked MEG data [126]. In general the potential approaches are two fold: Mapping existing brain data to ANN model activity (or using ANNs to map respective activity) or building brain data inspired neural network models from ground up inspired by brain functions and test functional principles thereby.

Experiment 1: brain machine SIMILARITY In a first step, previous attempts on mapping neuronal network activity to brain data and vice versa, need to be extended to the time-frequency domain. The study by Seeliger et al. (2021) [124] used fMRI data to train the CNN model. This model learned to produce the overall fMRI activation for several primary regions from the original input using 118.000 volumes of fMRI data as training examples. Instead of fMRI data, time-frequency resolved MEG could serve as input as well. Trial-by-trial fluctuations - depending on the input stimulus - would hence not be voxel activity measured as BOLD activity, but perhaps γ band power. Even more loosely interpreted, the proposed brain-ANN correspondence with respect to fundamental image processing strategies of both networks (if existing), should be visible even without "cross-training" the network with brain imaging data. Instead the hypothesized functional similarity should be possible to extract even from a pre-trained CNN model, such as the VGG16. This network has been trained to classify 1000 image categories which is probably less than a human is trained on, but still reasonable in size. If - as hypothesized - CNNs and the brain share computational similarities and those similarities are biologically reflected in temporal rather than spatial correlations (as with the CNN), then relating the frequency spectrum of human participants in primary visual regions - exposed to certain stimuli - to the activation of more shallow e.g. VGG16 network layers - exposed to the same stimuli - should reveal some correspondence. One possibility would be to use the MEG data that has been collected during Experiment 2 (see page 130) and related primary visual region time-frequency resolved data to the activation pattern of the VGG16 when exposed to the same stimuli. To approach such an experiment, the raw MEG would need to be transformed into source and time-frequency space, e.g. by using a DICS beamformer [423] or time-frequency transformed virtual channels obtained from LCMV beamforming, such that each grid point within the gray matter of the subject's e.g. V1 is transformed into a separate channel. The resolution of the fMRI data used by Seeliger et al. (2021) [124] was relatively low (2.4 mm isometric voxel size) and hence a grid spacing of around 2.5 mm for the beamformer approach might already yield sufficient results. One major problem however is the amount of training data. During MEG Experiment 2, 1.400 trials have been collected from every subject irrespective of the condition. Since 40 participants have been recorded using this task, theoretically 56.000 trials could be obtained which would result in high statistical power with respect to a given correspondence metric. If γ band power is related to feed-forward stimulus processing, then trial-by-trial fluctua-

tions in the γ band as a response to the stimulus presentation should be spatio-temporally related to the CNN activation patterns. Similar to what has been demonstrated by Seeliger et al. (2021) [124], a neural network could furthermore be trained (based on 56.000 trials) to produce respective time-frequency resolved data for each grid point. The learned spatial filters from this process are expected to be highly similar to what has been found for fMRI obtained filters. This is due to the high correspondence between γ and the BOLD signal [238, 303, 361], but furthermore due to the suspected high functional correspondence. If such a similarity could be demonstrated (e.g. for γ), then indeed the computational role of γ becomes directly accessible to some extent.

Experiment 2: brain machine MODELLING Furthermore, hypothesized computational functions for different frequency bands could be investigated *directly* on respective neural network implementations. One straight forward approach would be to utilize a pre-trained neural network, such as the VGG16 and transfer the respective CNN layer weights onto a spiking neural network (SNN) [438]. Resulting spiking neural networks expose - strongly dependent on parameter settings [104] - a variety of temporal dynamics. Those could be compared directly to brain data via neural mass models [439] or DCM simulations (see e.g. *The Virtual Brain* [112]³⁷). On a much simpler scale however, the behavior of the network could be assessed under the influence of a coupled external oscillator. It could be shown that SNNs can be "entrained" using a spiking neural oscillator (SNO) [440]. External oscillatory entrained activity would thereby correspond to the role of coherent α band activity in the cortex [91]. A SNN derived from e.g. a pre-trained version of the VGG16 via DNN-SNN layer conversion [438] could be connected to a SNO of low frequency (relative to the networks inherent time scaling) [441]. Classification performance *before* and *after* adding the oscillator potentially adds insight into the potential functional role of low frequency oscillations in the brain. It has been hypothesized that coherent low frequency α activity sets up a communication channel in the network, aiding the respective computations [91]. Hence, a coherent low frequency oscillation across multiple SNN layers would - to some extent - synchronize the internal network activation as well, which is hypothesized to aide certain (low coherent α or θ power) computational pathways [260], whereas others are suppressed (high α power). Since low frequency oscillations have been strongly associated with feedback activity [52, 82, 91, 420] and the classical version of the VGG16 does not have such connections, the transferred architecture potentially needs not be complemented. However, even without an implementation of the connections themselves, classical α inhibition effects [224] whether true or not [176] could at least be implemented on the level of the hypothesized influence on each layer, whereby the network's response (either behaviorally in terms of e.g. accuracy or computationally) is measured.

³⁷<https://www.thevirtualbrain.org/tvb/zwei>

Open research and discussion

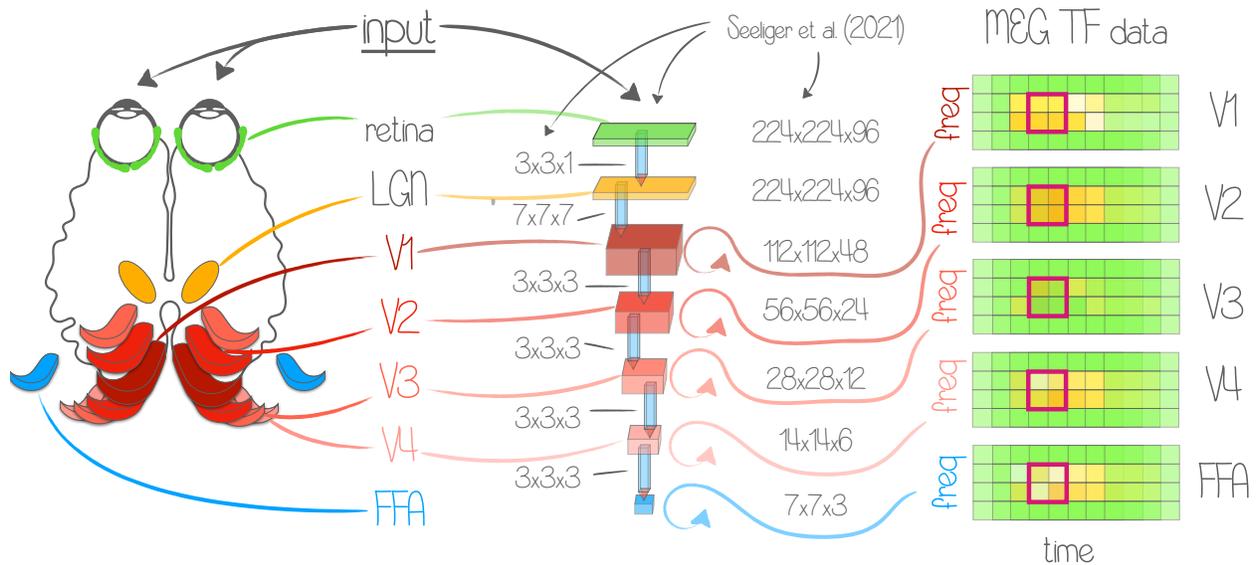


Figure 34: **Concept for ANN Experiment 1.** Derived from the human visual architecture, a neural network is set up to comprise, retina, LGN, V1-4 and FFA, similar to what has been done by Seeliger et al. (2021) [124]. Here however, V4 was added and the respective input size was doubled, while the rate of convergence was kept constant, leading to a same sized output. Instead of training the network on a single human subject of whom 23 *h* of fMRI data has been recorded, the model will be trained on data from time-frequency resolved virtual channels at respective grid points. The network will be trained to produce respective time frequency transformed MEG data *from* the stimulus. Afterwards, layer weights of the ANN are analyzed and visualized. Data from the second MEG experiment (see page 130) will be used for this ANN experiment.

Experiment 1: brain machine SIMILARITY Brain-machine-similarity can be measured multiple ways. However, two main strategies have been demonstrated to provide promising results: Relating brain and neural network activity directly, given a similar task [126, 289], or by training neural network models on brain data and extracting learned features [124, 442]. For both analyses, previous literature provides template analysis pipelines. From there hypothesized working principles could be derived. Seeliger et al. (2021) [124] already propose an architecture where a neural network model was trained to produce fMRI data from input stimuli that were similar to what the subject has seen during data acquisition. The network was set up such that primary visual regions (retina, LGN, V1-3, MT, FFA) are resembled. Due to the close relationship between the BOLD signal and γ band oscillations for the processing of visual stimuli [238, 303, 361], the principle experiment can be repeated, by replacing fMRI data with grid spaced single trial data bins of time-frequency resolved MEG data in the γ band for each virtual channel. Since multiple participants are used, respective

individual anatomies (an individual virtual channel locations) have to be transformed into e.g. MNI space [425] in order to achieve the required trial count. Here, the data from MEG Experiment 2 (see page 130) to train the network. Thereby, 40×1400 trials have been recorded over all participants. This yields 56,000 training examples (counting in all trials and ignoring the fact that probably some trials need to be excluded). For each trial - or training example - time-frequency resolved virtual channels are computed either using a DICS beamformer [423], or time-resolved virtual channels obtained via a LCMV beamformer [337] that are transformed into time-frequency data afterwards (see e.g. "Feature specific neuronal oscillations in cortical layers (in prep)"). The anatomical resolution is obtained from the recorded T1 weighted MRI, where each respective gray matter voxel is assigned to a given ROI and used as a grid point for the beamformer search. Not only could respective time-frequency bins of interest (e.g. shortly after stimulus onset until shortly before the response) be averaged and used as data points *exactly* like Seeliger et al. (2021) [124] did, but furthermore, a sliding window approach could be used, as has been done previously as well for comparing time-locked MEG data with CNN activation [126, 289]. In both cases, spatial filters learned by the model can be extracted and visualized. It is hypothesized that those extracted features expose structural similarities to what has been found by Seeliger et al. (2021) [124] for fMRI data and that those similarities are on average strongest where the average γ band response in that respective region is strongest. Furthermore, an exploratory analysis exploiting the same methodology is planned for θ , α and β as well. See Figure 34 for an overview of how the experiment is conceptualized.

Experiment 2: brain machine MODELLING In order to build a neural network model that exposes known similarities to the brain and can be probed to measure the network's response to external oscillations, a set of obstacles until then has to be overcome. *First*, a potential SNN architecture has to be found that a) can perform the task and b) shares structural (and temporal) similarities with the brain. *Second*, SNN neuron and network parameters have to be found that allow for the expression of realistic low (e.g. α) and high (e.g. γ) oscillations. Only then a respective perturbation to the system could potentially generate new predictions that can be tested using brain data. *Third* respective experiments have to be carried out.

In general, this part is the least conceptualized of all aspects of the recent work. It has been thought the most difficult task with the most pre-requisites (e.g. MEG data and GPU cluster access) and most unclear methodological parameters. Only a few pre-tests with respect to SNN networks have been conducted. In fact only the *second* part has been partly addressed. In a first step, LIF neurons were investigated with respect to their capabilities to express rich enough dynamics on the network level (α , γ). A SNN was implemented as proposed by Gu et al. (2019) [443] only with twice as many neurons (see

Neuron parameters			Network parameters		
C_m	1.0	membrane capacitance in pF	N_{exc}	800	excitatory neurons
E_L	0	resting membrane potential in mV	N_{inh}	200	inhibitory neurons
I_e	0	external current in mA	w_{exc}	0.1	excitatory weight
V_m	0	membrane potential in mV	w_{inh}	0.5	inhibitory weight
V_{reset}	10.0	reset membrane potential in mV	d	5	synaptic delay in ms
V_{th}	20.0	firing threshold in mV			
t_{ref}	2.0	refractory period in ms			
τ_m	20.0	membrane time constant in ms			

Table 3: **Network parameters** of LIF model to estimate γ band response depending on stimulus gain, reproducing previous results [443]. See Figure 35 bottom.

most relevant parameter settings in Table 3). The network architecture has been tuned towards edge detection in a wider sense. 800 excitatory and 200 inhibitory neurons have been connected all-to-all mimicking a grid of ON or OFF cells (see "Structural aspects of the brain"). Depending on the strength of the synaptic input (Poisson distributed input spikes to each neuron), the neuronal dynamics change. Findings by Gu et al. (2019) [443] could be replicated (see Figure 35 bottom). If the input gain was set to exactly match the average response threshold of the network, the network predominantly expressed 10 Hz oscillations (Figure 35 bottom left). Doubling, tripling or quadrupling the input gain, shifted the peak frequency in the network from $\approx 70 Hz$ to $\approx 105 Hz$ to $\approx 135 Hz$. Note, that only the amplitude but not the frequency of the stochastic input process has been altered.

In a second experiment, the same number of all-to-all connected neurons have been chosen only this time implementing the Izhikevich neuron model [104]. The respective parameter settings can be obtained from Table 4. A constant input of duration T has been provided (similar to the LIF model) however each neuron was tuned to a different input current. This simulates the response to e.g. differently oriented bars, where neighboring stimulus features elicit weaker but still substantial responses [30]. A response time of $D = 50 \times T$ was measured. Setting $T = 20 ms$ results in a trial length D of 1 s . Figure 35 top middle and right show the spike raster plots of each single neuron, whereas Figure 35 depicts the time frequency transformed average membrane potential V_m over all neurons. As a response to typical single feature stimulation, the inhibitory neurons within the network synchronized in a low frequency rhythm ($\approx 10 Hz$) which entrained the high frequency excitatory neurons that otherwise would expose increased activity between 100 Hz and 150 Hz which is comparable to what was found for the LIF simulation.

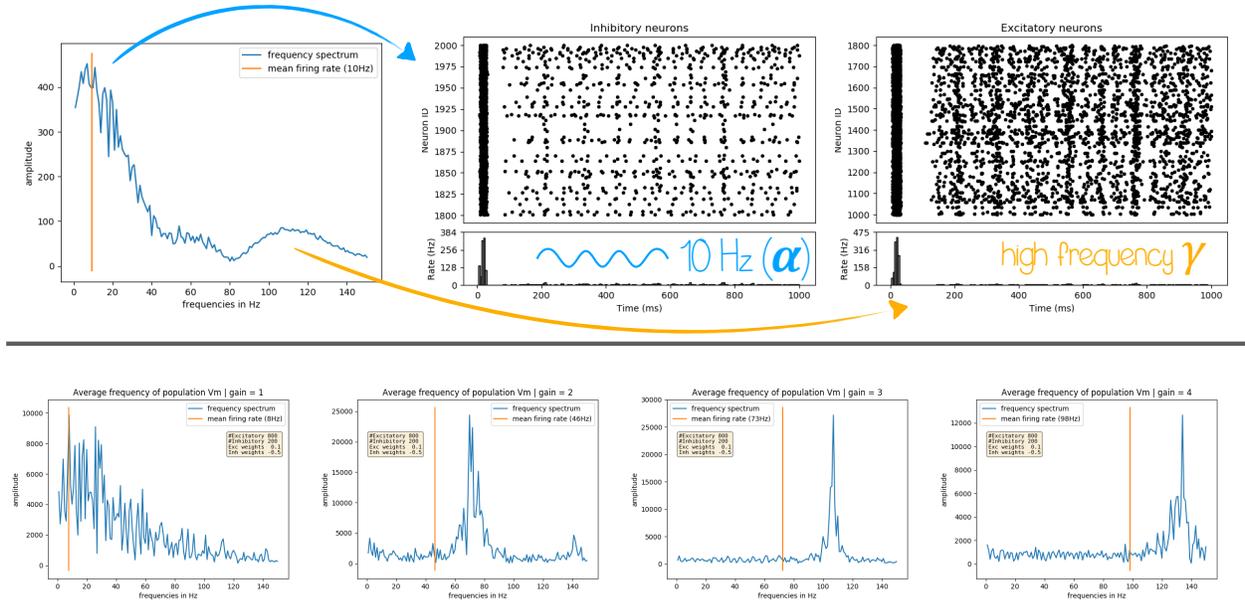


Figure 35: **SNN to stimulate low and frequency states and SNN response to grating stimuli**
The *upper part* of the figure was derived from a neural network simulation based on the Izhikevich model [104] to stimulate inhibitory low frequency oscillations that are reflected in excitatory responses as well. See Table 4 for the set of parameters. The left side reflects the frequency spectrum of the active neuronal network. Middle and right parts depict the spiking events of inhibitory and excitatory populations within the network. The network was tuned to expose inhibitory α oscillations on receiving input. As clearly visible in the frequency spectrum (left) and the spike raster plot (right) low frequency oscillations from inhibitory neurons impose rhythmicity on excitatory neurons as well, but including a more or less synchronized high frequency component (here around 90 Hz to 150 Hz). The *lower part* depicts the time frequency spectrum of a different SNN based on LIF neurons [443]. This neurons were tuned to respond with γ oscillations to input grating stimuli (alternating brightness values). Thereby the gain ("contrast") of the stimulus has been modified, which caused a shift in the frequency spectrum, of the γ band.

Neuron parameters			Network parameters		
N_{exc}	800	excitatory neurons	a	$0.02 + 0.08\alpha$	time scale of recovery
N_{inh}	200	inhibitory neurons	b	$0.25 - 0.05\alpha$	sensitivity of recovery
V_{th}	30	firing threshold in mV	c	$-65 + 15\alpha^2$	reset value of V_m
V_m	-65	membrane potential in mV	d	$8 - 6\alpha^2$	reset value of U_m
V_{min}	-65	lower bound in mV	α	$X \sim \mathcal{U}(0, 1)$	random between 0 and 1
			U_m	$-65b$	V_m recovery

Table 4: **Network parameters** for the Izhikevich model [104] to simulate event related α and γ band responses. See Figure 35 top.

Future considerations

Both experimental strategies have been explored theoretically but could not finally be developed to a degree where a first feasibility assessment could be made. Many methodological questions still remain open and subject to further investigation. A first success on the road-map to completing Experiment 2 (see page 155) has been to replicate findings by previous research on the simulation of low and high frequency oscillations using SNNs [104, 443].

Experiment 1 (see page 154) has not exceeded conceptualization phase. From the now gathered MEG data the network can be trained as described, based on promising approaches from previous literature [124, 126, 289]. It is expected that γ band oscillation trial-by-trial power can serve as a similar model as previously used fMRI based models. Indeed, if expected results became manifest, more evidence towards γ reflecting feature *type* specific *processing* [91, 189] rather than representing feature specific information or binding [184, 444] could be presented. For Experiment 2 (see page 155) at least a promising pre-selection for respective neuron models could be made. However, the trained model and hence respective experimental results are still lacking. In a first step, a pre-trained image recognition network - such as the VGG16 - could be converted into a SNN [438]. Parameters derived from the pre-experiments thereby serve as starting points for further tuning. By adding an external oscillatory source to the SNN while image recognition is performed, behavioral performance (i.e. accuracy) can be measured depending on the frequency, amplitude and phase of one or more simultaneously active sources. In addition to that the pre-trained VGG16 could be trained further to perform the task used in MEG Experiment 1 (see page 123). Afterwards the weight extraction and conversion would be performed similar to the original version. In one type of experiment, a bottleneck (i.e. limited access to the input) could be introduced and a controller added to the network, which allows the network to control its input data frame. This process is similar to what is thought how attention is implemented in humans [445]. It is hypothesized that simultaneous activation of input and output - depending on the location of the artificial attentional spotlight - leads to low frequency synchronization in respective neurons. In a second type of experiment those hypothesize low frequency oscillations could externally be imposed and the respective task performance (i.e. classification accuracy) is measured.

Irrespective of which approach is targeted first, the neural network development inherently relies on the MEG data analysis which as of today (January 15, 2023) has not been completed.

3 General Discussion

The human neo-cortex is of special interest as a computational model, due to its extended problem solving, motor coordination and linguistic skills. Hence, an understanding of the computational principles in the cortex does not only provide insight into ourselves from a philosophical perspective, but furthermore offers the opportunity to transfer respective findings to artificial machines. In other words, if the human brain is capable to organize input-output response patterns, such that a human shaped flexible body can autonomously ride a bike through big city traffic at an average power consumption of 20 *Watt/h* for the control unit [17], then a machine, based on brain-compute principles, would provide a major leap to the current transistor based ANNs which would consume *Gigawatt/h* of energy to achieve human level performance [18]. In addition to that, understanding the exact compute architecture of the brain provides new methodological perspective in the field of neuro-prosthetics and BCI. Thereby, understanding the *exact* neuronal code might be deemed impossible for the foreseeable future, since even the neuronal code of the *Hydra* which only has a few hundreds to a few thousands of neurons which all have been recorded simultaneously, could not be deciphered [24]. This "black-box" problematic of the brain is shared with ANNs.

3.1 Summary of results

We could demonstrate that α and γ oscillations could be mapped differentially to separate cortical layers within V1, but furthermore that α and γ band signals are differential related to feature processing as well. As predicted by previous literature [52, 91, 238, 272], γ band oscillations have been found to be positively related predominantly to deep and superficial layer BOLD activity and has been linked to ongoing feature processing (higher γ power for **Pc** – **nPc**). γ oscillations are thought to reflect feed-forward processing [52, 91, 240]. To confirm this hypothesis it would be necessary to conduct a connectivity analysis similar to previous literature [265], which remains to be done. In general, predicted findings for the γ band have been confirmed [52, 82, 272]. However, the data suggests that γ does not directly reflect *specific* feature processing but probably rather *general* feature processing. On a side note, the here observed EEG γ band responses (see Figure 15 (B)) are comparable to what has been found for the preliminary analysis of the MEG data of Experiment 1 (see Experiment 1 on page 123 and Figure 30). A positive γ band response could be observed post-stimulus and early trial (200 *ms* to 400 *ms* after stimulus onset) for the combined activation of *any* stimulus. This verifies that at least expected fundamental γ components are present [173]. Instead of direct feature selective processing γ oscillations have been thought to reflect a coherent feed-forward signal of prediction errors [422]. Since in the presented EEG-fMRI experiment the orientation of the stimulus did not expose any informative value for the participants with respect to the task and since the actual features of interest have been

presented randomized in an odd-ball paradigm, presented findings are not easy to reconcile with this framework. This is different for the MEG data set of Experiment 1 (see Experiment 1 on page 123), where specifically predictability has been modulated. As the MEG data remains to be investigated, no conclusions about the relationship between γ oscillations and stimulus predictability can be made. However, first behavioral results indicate that even though participants were not even able to explicitly report the presence of some probability scheme (let alone the scheme itself), RTs of easy trials monotonically increase with decreased predictability. This indicates that predictability did indeed affect the stimulus processing to some extent. A more detailed interpretation of current findings with respect to current theoretical frameworks is presented below.

α band activity has been found to be negatively related to the BOLD signal between 8 Hz and 14 Hz with strongest negative effects in deep and superficial layers for activated (relative BOLD increase) and in superficial and middle layers for deactivated voxel (relative BOLD decrease), irrespective of the response specificity (amount of voxel included in the contrast). Negative correlations between α power changes and the BOLD signal in superficial layers have been found to be related to more general activation / deactivation processes driven by directed (spatial) attention [56, 270, 354]. First results obtained from the pre-stimulus period (500 ms prior to stimulus onset until stimulus onset) - computed by Maryam Mostafalu [1] - indicate the presence of a lateral attention effect over occipital sensor sites in the α band, in line with findings from previous literature [173]. This means that potentially a targeted (dis-) inhibition could be observed (hemisphere specific) for the respective given attention condition. Since this effect occurred pre-stimulus, feedback directed activity is likely to be the cause. Note, that no comparable results have been found for the γ band, which serves as a sanity check. No laminar specificity has been investigated yet, however a general attention lateralization selective process has been observed similar to the general (de-) activation process observed for the EEG-fMRI experiment. While this effect is consistent between preferred and not preferred orientation for deactivated voxel, the general α response pattern differs when only considering generally activated voxel. It is hypothesized, that an interaction between general attention and feature selective processing results in a non-additive relationship between attention α and feature α for the preferred over not preferred orientation. Interestingly, a feature and frequency specific difference could be observed for oscillatory responses in the α band. Comparing the correlation between voxel preferring one feature with EEG α with voxel preferring the other feature, revealed a differential response pattern for low (8 – 10 Hz) and high (11 – 13 Hz) α . Thereby, fMRI BOLD activity decreases for both, voxel preferring a respective orientation (**Pc**), which is correlated negatively with upper α frequencies and voxel preferring the respective other orientation (**nPc**), which is reflected in lower α frequency. Hence, upper α oscillations are related to increased processing of the preferred stimulus and decreased processing of the not preferred stimulus and the reverse for lower α frequencies. This predominantly deep and

middle layer α effect is present irrespective of the respective attention sub-selection (A^\pm). Strikingly, even though no significant difference between \mathbf{Pc} and \mathbf{nPc} has been found with respect to the α frequency profiles, superficial layer profiles for \mathbf{A}^- and \mathbf{A}^+ for \mathbf{nPc} appear rather similar. On the other hand, the profile for \mathbf{Pc} deviates strongly from \mathbf{A}^+ compared to \mathbf{A}^- , such that an interaction between α activity related to the general activation of superficial layers and feature selective α activity of deep layers could be suspected. This hypothesis receives further support from the fact that we found attention related deactivations in the BOLD signal for feature and feature signal-free voxel selections in superficial layers, linked to α . The canonical microcircuit model of the cortex (see "Canonical microcircuits: cortical compute modules") [82] predicted low and high frequencies to be related to deep and superficial layers. Low frequency (e.g. α or β) oscillations have been attributed mainly to feedback processes [52, 82, 91] in superficial and deep layers [52, 80, 272].

In Experiment 2, where more long range connections between lower order and higher order cortical regions have been targeted for investigation preliminary results have been kindly provided by Maxime Ferez [407]. In a face-word Stroop experiment either facial or lexical features of a compound stimulus of both needed to be attended. It has been hypothesized that α band activity coherently sets up the communication channels between specific feature processing neurons in lower and higher order regions and inhibits rivaling information [91]. A conflict between information derived from facial features with information derived from the semi-transparent overlaid word should be visible differential by a relative α power decrease over target (FFA) and a relative α power increase over distractor (VWFA) processing regions. An analysis of the pre-stimulus period of Experiment 2 revealed significantly higher α band activity for the attend-face condition compared to the attend-name condition over cortical areas that include VWFA. Furthermore, an analysis of the attend-face condition alone revealed that increased α over VWFA regions was stronger for in-congruent trials, where the RT was lower. This indicates that successful distractor inhibition improves behavioral performance [432]. Since it appears highly questionable whether distraction *inhibition* indeed causes behavioral improvements [176], the presented results could furthermore be interpreted with respect to modified sampling of information [180, 446].

3.2 Interpretation in the light of current theoretical frameworks

Here presented theoretical models and experimental findings help to shed light onto those core functional computations. Findings from the laminar EEG-fMRI experiment not only replicate findings from Scheeringa et al. (2016) [238], but furthermore provide novel insight into the laminar frequency architecture of the human visual cortex. Multiple sources for synchronized neuronal activity in a frequency range - often called α - has been suggested depending on the task and respective cortical region [234]. Bonnefond et al. (2017) [91] propose the presence of at least two functionally distinct processes that are implemented via

α frequencies: setting up a neuronal communication channel for feature specific processing (superficial layer) and a generally modulating activity mainly in deep layers. Since active feature processing and attentive processes are very likely interlinked (participants have been explicitly asked to attend and process), activation increases reflected by a relative increase in BOLD response, cannot be decomposed into feature and attention components easily. However, given that the main contrast (preferred - not preferred) did at least implicitly control for (spatial) attention, this contrast might be mainly driven by feature specific processes. It has been found that attention has a major influence on the BOLD signal in V1 [447] and can modulate fMRI responses in a spatially specific manner [448]. Furthermore, it has been shown that attention modulates the EEG signal particularly in the α band [251]. If selective attention is related to positive changes in the BOLD signal, then investigating voxel responding with a relative negative change to the presentation of a stimulus might reveal feature specific suppression. Indeed, a negative correlation between EEG α and the BOLD signal for all activated voxel could be observed for frequencies between 10 Hz and 13 Hz predominantly in the middle layers. When limiting the analysis to deactivated voxel (any stimulus compared to baseline), the negative correlation is preserved. However, the strongest effects are found in superficial layers and are more broadband. Regressing out the feature specific signal from the general activation and repeating the deactivation analysis did not alter those findings. Hence, superficial layer α band correlations contribute most to a negatively deflected BOLD response (relative to baseline) irrespective of the stimulus feature. Neurons processing features that are not preferred given a respective stimulus expose a negative correlation with α , but a lower BOLD signal in general. Higher low frequency α hence indicates a suppressive effect, similar to hemifield suppression in lateral attention tasks [251]. Attention modulated α (dis-) inhibition in general has been hypothesized a general compute principle of the cortex implementing the gating of cortical information flow [91, 148, 224] (see also "Gating by inhibition"). In addition to superficial layer attention related α we identified a deep layer feature specific component where upper α frequencies are more negatively related to the BOLD signal for preferred as compared to not preferred features, whereas the reverse has been observed for not preferred features in the lower α frequency band over preferred (see Figure 19). The difference between lower and upper frequency α in the overall frequency profile indeed indicates feature and frequency selective processes in deep layers. Note however, that a differential response in superficial layers for activated voxel in **Pc** compared to deactivated is suspected, but has not yet been investigated (see Figure 20). It is expected that superficial attention α and upper frequency deep layer preferred feature α interact only when solely generally activated voxel are selected. Findings furthermore provide evidence for a change in α frequency relative to whether stimuli needed to be integrated or segregated with respect to the temporal domain [449]. For a task requiring temporal segregation α frequency increased contra-lateral to the side of attention and decreased ipsi-lateral [181]. Respective visual hemifields have been associated with target and distractor stimuli. The present findings on frequency specificity from the laminar EEG-fMRI experiment could

reflect a similar process, where not temporal features but instead visual feature processing is separated depending on relevance (preferred compared to not preferred orientation). A plausible explanation for the present data would hence be a model where incoming activity (received at L4) is modulated by feedback activity in deep layers (from higher order regions) [52], such that the respective signal is phase coherent with the higher order area in the α band which would allow the exchange of information [91]. Employing the procedure of nesting the incoming γ band signal into the feedback driven α in deep layers would be a way to recruit a certain set of feature specific neurons. If neuronal populations are glued together by coherent low α power [91], then neuronal suppression cannot be in the same frequency range, if processing populations are spatially very close (like for linear features) in V1, because of the interference with the recruitment α . Recent theories have emphasized this role of α oscillations in information sampling [180, 446]. Since feature selective processing (deep and superficial layers) is inevitably mixed with signals related to "general attention", but attention related signal changes were mostly attributed to superficial and middle layers, deep layer correlations are assumed to reflect "pure" feature specific α activity. Given the framework by Bonnefond et al. (2017) [91], higher order regions are thought to recruit lower order regions using coherent α oscillations. At this stage, our data speaks in favor of a feature selective process, where feature separation is achieved by multiple α sub-bands [181] mostly in deep layers [52]. The canonical microcircuit model by Bastos et al. (2012) [82] furthermore assumes feedback related connectivity from deep layers to L4 within a single processing unit (e.g. column). In the light of the nested oscillations framework by Bonnefond et al. (2017) [91], incoming activity at L4 either from thalamus or lower order cortical regions [80] is thereby entrained with feedback α activity (here via deep layers). Since the targeted entrainment for preferred features is mainly found in higher α bands (11 Hz to 14 Hz), a potential overlap with β (typically starting at around 12 Hz) cannot be excluded. The predictive coding framework (see WHAT IS ... on page 33) assumes top down predictions to be reflected by feedback β changes. However, it has been hypothesized that those changes are conveyed by α reflecting precision [219]. A targeted recruitment of feature selective neurons implemented by high frequency α / β would be conceptually close to a "prediction". From the expected sensory content, respective feature selective neurons might be chosen in order to facilitate bottom up processes. This process might already start prior to the stimulus [252], but after stimulus presentation the respective predicted percept could be locked into the evidence based state [179]. The prediction error could be used as specific marker on how to update the previous set of allocated low hierarchy neurons, such that a respective value functions is maximized (e.g. feature correspondence) or the current state of processing is stabilized to collect evidence [181]. Bonnefond et al. (2017) [91] hypothesized feature specific α band activity to be reflected in superficial layers, which is not challenged by present results. Feature specific activity inevitably passes superficial layers in order to be forwarded to the next processing stage [80] and the EEG-fMRI experiment demonstrated those to be related to feature processing as well.

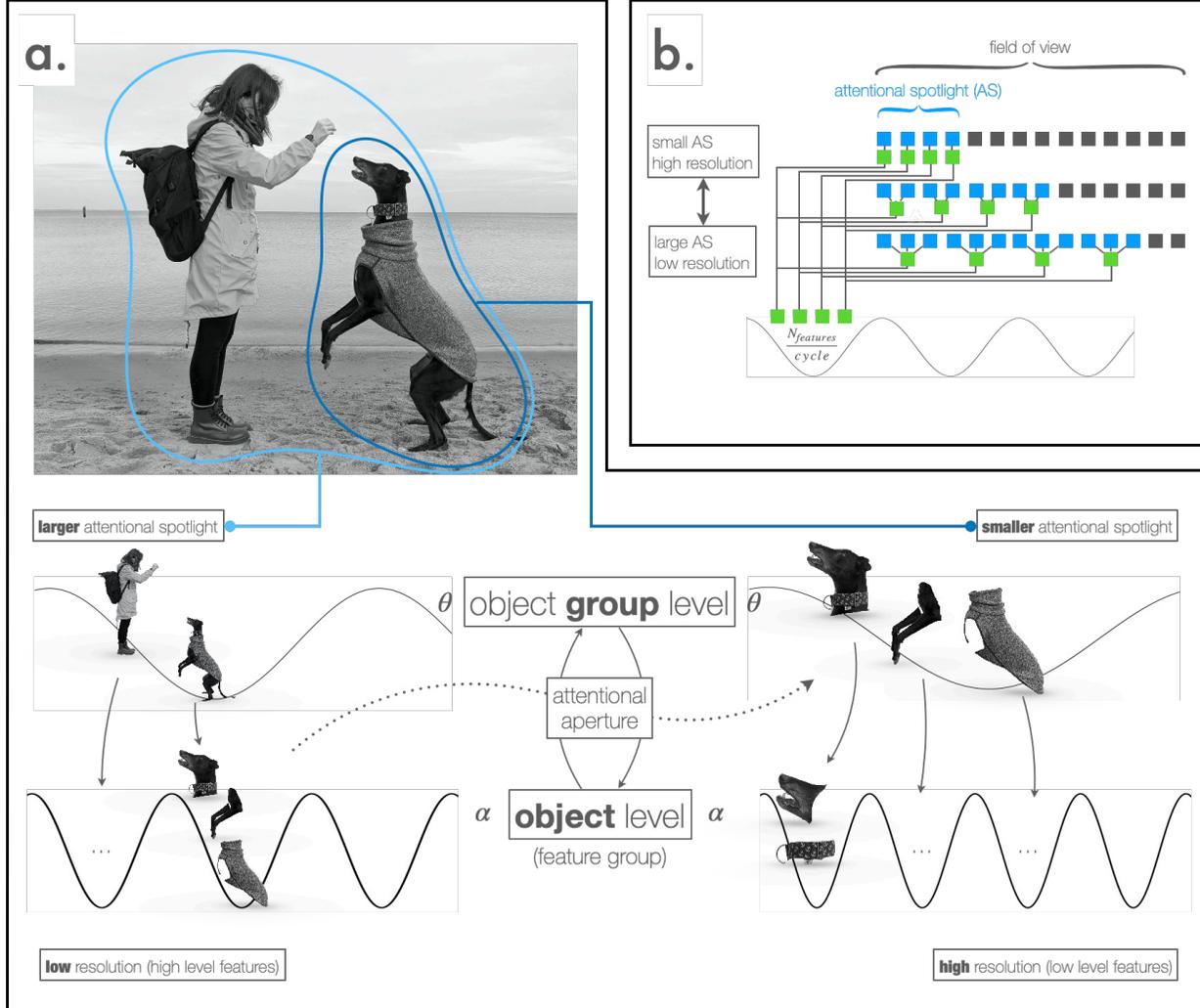


Figure 36: **Scalability of phase dependent object and feature processing.** A) Object groups are redefined depending on the size of the attentional aperture (assuming a constant field of view). Attending the woman and the dog might form an object group of two objects (the woman and the dog). Both objects are sampled at different phases of low frequency oscillations (θ). Features of each object will be processed in separate α cycles per object, along the phase gradient. Thereby salient, goal driven or otherwise more important features will be sampled early in the cycle. Once an “attentional aperture change” is applied (i.e. the attentional spotlight is shifted to the dog only) a rearrangement would occur. A new object group is formed, consisting of more low level “sub-objects”. Those sub-objects are again sampled at θ frequency. This allows for a higher feature resolution at the (sub-) object level. Features are again sampled / processed within separate α cycles. This inverse relationship between size of attentional aperture and resolution within the attentional spotlight could be demonstrated by [450]. B) Due to the limited processing capacity within a single α cycle, low level features need to be grouped into higher level features, such that the entire object processing still fits within a single cycle, which results in a decrease of spatial resolution for a wide attentional aperture. Processing visual input at the highest possible resolution (e.g. to determine the exact shape and texturing of the dog’s collar in A), could only be performed on a small fraction of the visual field. The processing of the entire scene in turn requires features to be “downsampled” and / or grouped in order to fit a single cycle length. This figure has been reproduced with kind permission of Bonnefond et al. [258].

Once the incoming signal (at L4) is entrained via deep layer α connections [52, 80] it is routed to superficial layers, where the feed-forward transmission mostly happens [80, 82]. I would propose a second source of α which seems mostly related to signal gain modulation that decides via its inhibitory strength whether the incoming signal, convolved with deep layer α activity, can actually pass to the next higher order region (L4). No frequency specificity could be observed for negative correlations between EEG α and BOLD signal in superficial layers indicating a general process as well. Hence, multiple simultaneous processes reflected in the α band could be dedicated to different tasks: general (probably spatial) attention in superficial layers [270] and feature selective "module selection" via deep layers [52]. I would hence suggest that by predicting which cortical modules are required and allocating them even prior to the stimulus (deep layers), higher order regions recruit a set of task specific neurons via deep layer α coherence. From deep layers, the α signal is probably routed to L4, where it picks up the "content" from lower order regions (reflected in γ) and nests them into its phase gradient. In superficial layers, a second α related process fulfills a gate keeping activity, similar to what the GBI hypothesis describes [224] (see "Gating by inhibition"). E.g. spatial attention would be assumed to act at this level, where the signal is hindered to leave at superficial layers. However, if superficial α is low (e.g. due to attention), information that has been brought into coherence with high order region α can travel to L4 of that region.

A relatively novel viewpoint has recently been brought forward by Bonnefond, Jensen and Clausner [258] in a not yet published article. Figure 36 has been reproduced from this paper to illustrate the core idea. This framework interprets previous literature in the light of a more sampling driven low frequency mechanism. In order to keep this section contained, the respective evidence will not be discussed here in detail. Instead only core assumptions or predictions that are derived from this model are presented and very shortly discussed. This was done to relate findings within the realm of this thesis, to novel - yet unpublished - ideas about ongoing cortical processes without having to present and justify my arguments in detail. It is argued that α and θ band phases are used to sample specific aspects of one or multiple objects. Thereby, neuronal activity is "sorted" along the α phase gradient to separate certain object *features*, which are nested in a lower frequency θ phase gradient which drives object sampling. An increase in amplitude of α and θ has been suggested to reflect narrowing down the "window of opportunity" (decreasing the number of phase points where inhibition is low enough for an action potential to be released) for neuronal signalling (increasing object or feature separation). A shift in frequency would be seen as an increase in the respective sampling rate. The described process would correspond to the feature and frequency specific α effect that has been found in the EEG-fMRI experiment for deep (and superficial) layers, since the proposed "sampling α " is hypothesized to be set up via feed-back directed predictive processes that could be attributed to deep layers [34]. Increased α power and decreased frequency would reduce the sampling capacity in voxel responding to the not preferred orientation, which effectively implements a suppression. The model further

hypothesizes that the relative size of the attentional spotlight, with respect to the size of a given object, indirectly modulates the feature (or object) resolution, by grouping more and more feature specific neuronal activation along the phase gradient. However, this respective aspect could not be tested here. Please see "My personal framework: The convolutional brain" where I propose my own interpretation of the data and previous literature in a speculative way. The model proposed in that section shares a lot of predictions with the here described sampling model.

3.3 Overall progress of the project

The present project aimed to add to the understanding of cortical computations on a local level and in widespread brain network dynamics. Oscillations are thought to play an important role for canonical microcircuit connections [80, 82] and long range higher order functional communication [253, 451]. It has been hypothesized that multiple attention and feature related α band processes modulate excitability *targeted* (e.g. stimulus specific) and *coherently* [91]. Here, I present first evidence for feature and frequency selective α band activity in deep and middle layers. Furthermore, superficial α power decreases could be shown to be related to the general level of excitability of large patches of cortex, potentially receptive field specific [48]. To the best of the authors knowledge, this demonstrates for the first time a differential feature, layer and frequency specific α band activity potentially reflecting different attention and feature related processes in humans using non-invasive methods. As has been previously demonstrated, the feasibility of laminar level EEG-fMRI to investigate cortical oscillations in humans can yield layer specific frequency profiles. Here the range of possibilities that this method enables has clearly been extended to the level of differential cortical processes. However, it remains to be investigated how cortical connectivity changes relate to α band power [265].

In addition, α - γ phase coupling remains to be investigated as well, which could not be done using EEG-fMRI on a laminar level. Two additional MEG experiments have been planned, designed, implemented and executed. One has been set up to target local microcircuit connectivity, whereas the other one was set up to target long range connections from higher order regions. Heavily impacted by the COVID 19 pandemic, recordings for both MEG experiments had been delayed for a substantial amount of time. Only the conceptualization and implementation of respective experiments and all subject recordings could hence be completed. Preliminary analyses, kindly provided by colleagues however, already allow for a first peak on the data. Both experiments had been set up to test different aspects of local or more widespread cortical activity in the visual domain. Experiment 1 (see Experiment 1 on page 123) specifically targets local computations in early visual cortex areas, whereas Experiment 2 (see Experiment 2 on page 130) tests hypotheses with respect to stimulus dependent long range connections between primary areas and higher order regions.

Data provided by colleagues [1, 407], already indicates that key predictions are met which can serve as a sanity check for ongoing analyses. Thereby, predicted occipital α lateralization has been observed for the lateral attention task and pre-stimulus α inhibition in higher order cortical areas for the face-name Stroop task. A post-stimulus generally stimulus driven γ response has been observed for the lateral attention task as well, but no other meaningful analyses with respect to γ have been conducted so far. Thus, respective ROI source level (both, classical and laminar [220, 268]) and connectivity analyses, necessary to collect evidence with respect to the proposed frameworks, remain to be done.

The setback caused by the recording delay was mildly compensated by enabling the author to invest time in different scientific disciplines. To solve a respective statistical problem - the determination and statistical testing of layer activation profiles for the laminar level EEG-fMRI experiment - common statistical procedures did not suffice. Research into the respective topic resulted in the aros test [353]. Via a permutation procedure it is tested, whether the ranks of more than three group averages are indeed justifiable by the data. It could be shown that - under certain circumstances - statistical power with respect to effect size is sacrificed in favor of a gain in qualitative differentiation power. For the EEG-fMRI experiments this means that the one or two most active layers can be determined and statistically backed, however the strength of this effect remains unknown. Nevertheless, the author does hope that this test will find application where classical test procedures either fail or result in too conservative test statistics.

With the completion of the MEG data recordings, the training of respective ANN models (see sections "Artificial neural networks") becomes feasible. In order to investigate brain-machine similarity, a CNN with a coarse "brain-like" architecture can be trained to reproduce the data collected in Experiment 2 (see Experiment 2 on page 130) from corresponding stimuli. While the raw data can be related to the network activity across multiple layers and time points directly [126], a time-frequency transformed source level γ band signal could be used instead of fMRI data to replicate findings from previous literature [124]. It is expected that due to the high similarity between the feature specific BOLD signal and γ band power changes [238, 303, 361], extracted learned filter kernels from the artificial "V1" expose patterns that, if convolved with the input stimulus, act edge enhancing. In a second step, either learned layer weights of the CNN or generic weights from e.g. the trained VGG16 would be transferred to a SNN architecture. First explorations of the parameter space of such networks, revealed low and high frequency stimulus-dependent responses [443]. It is expected, that early SNN layers express γ band activity that can be related to the active feature kernels, as well as α oscillations, synchronizing neurons relevant to the ongoing processing. α band oscillatory changes could also be implemented by external rhythmic stimulation, which has been found to support sampling based computations in SNNs [260]. In combination with recent findings, showing that brain-like feedback generators ("predictors") make networks

more robust against adversarial attacks [132], this suggests that implementing brain-inspired features to ANNs can improve their overall performance and advocates once more for an exchange of knowledge between the fields of neuroscience and AI.

3.4 Closing remark

Plants do not have brains or nervous systems [2]. As has been described mainly for the visual system, neuronal (cortical) computations are tightly linked to physical stimulus representations. Some researchers suggested that neuronal oscillations - or in fact their exact frequency - are directly linked to the environmental interaction (behavior) [146] or physical processes of the body [452] and *their* respective time scales. Neuronal oscillations are either the result or even required by some general principle of neuronal processing. For instance, *Honey Bees* expose a similar response pattern to different kinds of odor stimuli as would be expected for classical mammalian α band responses, but in the ≈ 18 Hz range [159]. It can hence be assumed that, since plants do not expose short lived *behavior* (few exceptions like the *Venus flytrap* aside), a corresponding short lived *perception* and corresponding neuronal coordination is not an inevitable requirement to prevent extinction. The assumption that functional principles of the human neo-cortex have been developed as a consequence of a change in nutritional supply and other evolutionary factor has received great support [453]. Combined, those findings are indicative for a computational setup in the mammalian (especially human) cortex that has been tuned towards energy efficient solving of problems related to the living conditions of a given species. I believe, the truth about how the brain works lies somewhere at the intersection of neuroscience and brain-inspired AI research.

In the last section (see page 173) I would like to role out a more personal theoretical framework that links artificial neural networks and neuro-scientific evidence. On a fundamental level modern artificial neural networks, such as CNNs have been derived from findings about cortical compute principles [73, 134]. A respective inverse linkage has however only recently been considered. The author believes that, since the brain is a neural network, the neuronal code itself might be inaccessible, however the respective functions that are computed could be interpreted as spatio-temporal filters, similar to a CNN and could potentially be extracted. Furthermore, accepting the assumption that artificial and biological neuronal networks share at least some computational similarities - which has been highly suggested [124] - enables neuro-scientists to derive testable predictions from artificial intelligence research. If there is such a thing as a more likely or less likely implementation for a given computation in *any* neuronal network, then artificial intelligence research and biological evolution might have converged at some points that need to be uncovered.

3.5 Evaluation of Hypotheses

Hypothesis 1 α power decreases over lower and higher order regions (or neuronal populations) are specifically linked to ongoing stimulus processing.

Evaluation: α oscillations decrease over lower order cortical regions are linked to active stimulus processing (see Feature specific neuronal oscillations in cortical layers (in prep) on page 58). For higher order regions the respective analyses remain to be conducted (see Experiment 2: feature attention and conflicting information on page 143).

Hypothesis 2 α oscillations are in synchrony (coherent) between lower and higher order regions (or neuronal populations) that are involved in ongoing stimulus processing.

Evaluation: The respective hypothesis remains to be investigated (see Experiment 2: feature attention and conflicting information on page 143).

Hypothesis 3 α power related changes with respect to attention differ from those with respect to ongoing feature processing, which would be expected based on anatomical findings [80] and the idea that those are implemented in separate cortical processes with different laminar profiles [91].

Evaluation: Results from the EEG-fMRI experiment (see Feature specific neuronal oscillations in cortical layers (in prep) on page 58) point towards a confirmation of this hypothesis. However, the results from MEG experiment 1 and 2 (see MEG experiments on page 117) need to confirm this. The respective analyses remain to be conducted.

Hypothesis 4 α power decreases are linked to increased activity in neuronal populations that preferred certain stimulus features as compared to the not preferred stimulus features.

Evaluation: The hypothesis could be confirmed (see Results on page 80).

Hypothesis 5 α band activity decreases for expected (predictable) stimuli [269].

Evaluation: The respective hypothesis remains to be investigated (see Experiment 1: spatial attention and predictability on page 134).

Hypothesis 6 α power changes are mainly linked to deep (possibly feature processing [34, 52]) and superficial layer (possibly spatial or directed attention [56, 270, 271]) activity.

Evaluation: The hypothesis could be confirmed (see Results on page 80).

Hypothesis 7 γ band activity is related to stimulus feature specific processes.

Evaluation: The hypothesis could partly be confirmed (see Results on page 80). However, γ band oscillations appear more related to the feature *processing* rather than the stimulus *features* themselves.

Hypothesis 8 γ band activity increases when predictions about stimulus features are violated.

Evaluation: The respective hypothesis remains to be investigated (see Experiment 1: spatial attention and predictability on page 134).

Hypothesis 9 γ band activity is related mostly to superficial and mid layer neuronal activity for ongoing stimulus processing [52, 91, 238].

Evaluation: The hypothesis could partly be confirmed (see Results on page 80). However, γ band oscillations appear more related to superficial and deep layers rather than superficial and middle layers, which is in line with previous research [52], anatomical findings [80] and current frameworks [91].

Hypothesis 10 γ band activity is nested in α band activity and is coherent between stimulus processing regions [240].

Evaluation: The respective hypothesis remains to be investigated (see MEG experiments on page 117).

Hypothesis 11 Spatial filters that transform an image into an edge enhanced version can be obtained by training a DNN to reproduce γ band power changes (obtained from a MEG experiment, see below) in the visual cortex from initial stimulus material.

Evaluation: The respective hypothesis remains to be investigated (see Experiment 1: brain machine SIMILARITY on page 156).

Hypothesis 12 External low frequency oscillatory activity increases robustness against noise in a SNN performing object classification.

Evaluation: The respective hypothesis remains to be investigated (see Experiment 2: brain machine MODELLING on page 157).

4 My personal framework: The convolutional brain

In this last section I would like to role out a - highly speculative - personal opinion on how *I* interpret the data in the light of previous literature. For this, I would like to establish two assumptions: 1) the brain *processes* information in the sense that it *transforms* (changes) an electro-chemically driven temporal code generated at sensory organs (e.g. retina) in a theoretically fully traceable manner. 2) the brain is a neural network that shares computational properties (e.g. such as each node or neuron producing a weighted sum from input activation or each neuron possessing a certain activation threshold) with ANNs. Especially the second assumption seems hard to justify. While the first one boils down to the observation of physical processes, the second implies a correspondence between biological and artificial neural networks. This correspondence is highly unlikely, since at least modern neural networks are built radically different, from what is assumed to be biologically plausible. However, while e.g. classical back-propagation used to train certain types of neural networks [84] are deemed biologically implausible [86, 87], a functionally similar but implementationally different mechanism might nevertheless be at play at dendritic connections [85]. Even though ANNs are knowingly implemented, the inner organization of how knowledge is acquired or represented is extremely difficult to access in a comprehensive way [454]. This is all the more true for the human brain especially since - as mentioned before - the *neuronal code* for simple organisms as the *Hydra* could not even be deciphered [25]. Even though the "machine language" of the brain remains a mystery to a large degree, respective functions that are implemented by this brain machine-code could be comparable to those of ANNs, at least in some domains and to some degree. Especially in the visual domain and early visual cortices, similarities between CNN activity and brain data are striking [30, 134, 455]. In order to understand *what* is computed in the brain, theoretical models about how a neuronal footprint given a certain function would look like must be developed. Since this pool of potential candidate functions is mere infinite, it could be narrowed down by hypothesizing that some functions the brain computes are actually similarly implemented in ANNs. To me, conceptual considerations, findings within the realm of this project and previous literature speak in favor of a computational concept, implemented in the visual system, that shares high similarities with CNNs. Functional requirements of a CNN and neuro-scientific evidence thereby provide the conceptual framework.

As described above, feature selective α in deep layers is hypothesized to reflect feature specific neuronal acquisition. A side effect of this process would be, that in the lower level region that receives the coherent feedback in the α band, all neuronal populations that are addressed, would be coherent in deep layers as well. Hence, information *within* feature specific neuronal populations that have been allocated to process a certain stimulus can be shared as well. In computer vision, image processing or CNN image classification a filter kernel is

used to enhance certain features of a stimulus. For instance a Sobel filter [456] $\begin{bmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$ can be used to transform a source image into a version with enhanced edges ³⁸ by simply convolving ³⁹ the input image step by step with the kernel. While the Sobel filter is custom made, CNNs typically learn the respective filter values from the data. The Neocognitro [134] can be seen as a predecessor to modern CNNs. It was based on findings on simple and complex cells in the visual cortex, responding with differential firing rate patterns to varying bar orientation or movement directions [30]. Complex cells are thought to combine simple cell responses similar to a convolution operation, where the stacked activation of multiple layers finally results in a joint "percept". Findings from the presented project - as well as previous literature - could be interpreted such that the targeted allocation through deep layer feedback α reflects the recruitment of neurons for a specific filter operation. Since the neuronal code is time dependent, α band activity at a specific target region would provide windows of opportunity, where neuronal signals can be transferred [91, 224]. Top down predictions would hence be realized by predicting the respective set of filter weights, necessary to maximize the value function, where α determines the "spatio-temporal width" of the filter by coherently recruiting neurons [91]. The " α - convolved" input signal is transferred to superficial layers from where it is routed to L4 of the next neuronal assembly [80]. In this model, γ oscillations would reflect the transformation of the signal. This would yield no feature specific result per se, but rather a "filter specific" γ (response to a specific kernel; basically the operation itself).

Recently, θ has been suggested to reflect attentional sampling of objects in the visual and auditory domain [457–459]. Thereby, θ would - similar to α - reflect some sort of input related sampling but at a different spatial scale [309]. It has been hypothesized that θ - γ coupling may reflect a possible implementation for coherence based long range communication [204]. The idea that θ coordinates object, whereas α coordinates feature sampling (possibly within a certain object) can partly be supported by the literature, since at least behavioral performance fluctuates in respective frequency bands [210]. Phase coherence in the θ between

38

$$\begin{bmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} * \text{img1} = G_x, \begin{bmatrix} +1 & 0 & -1 \\ +2 & 0 & -2 \\ +1 & 0 & -1 \end{bmatrix} * \text{img1} = G_y \rightarrow G = \sqrt{G_x^2 + G_y^2} \rightarrow \text{img2},$$

39

Convolution is a mathematical operation, where values of e.g. two similar sized square matrices (kernel size) are added and multiplied such that the resulting value for the center element is

$$\left(\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} * \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \right) = (i \cdot 1) + (h \cdot 2) + (g \cdot 3) + (f \cdot 4) + (e \cdot 5) + (d \cdot 6) + (c \cdot 7) + (b \cdot 8) + (a \cdot 9)$$

Hence, this kind of convolution can be seen as a (sliding window) weighted sum between a fraction of the input and a weight matrix (kernel).

V4 and PFC has furthermore been demonstrated to be a predictor for short term memory task performance [420] again indicating a role for item (or object) organization. Within the hypothesized model, grouping sets of features to objects via attention, would boil down to a targeted relative increase in gain. To group and potentially sort objects along the θ phase gradient, the window of opportunity for neuronal activity to be integrated within the same reference frame (one cycle) must be wider than the window of opportunity for feature phase sorting which requires a lower frequency compared to α . The respective synchronization would be implemented using a filter kernel as well. Learned values of a respective " θ -kernel" (jointly activated neurons in one regions within one θ phase) would modify the respective input image such that groups of jointly activated α coded feature maps (i.e. objects) are enhanced or suppressed. A respective analogon in classical ANNs would be the kernel size, where a larger kernel transforms more values into a single value (). However, due to the high inter-connectivity, respective kernels do not necessarily need to be spatial adjacent, as mere activational adjacency (in a given time window) would suffice. This means that beside spatial attention modulating relative neuronal gain, any other aspect of the input could be attended using the very same process. The filtered input would hence be the result of a joint activation of neurons selected from deep layer feedback α feature selection that is modulated via superficial attention related α and embedded in a general pool of activation that potentially spans large distances.

On a local level γ band oscillations would reflect the actual ongoing feature processing, or in the language of ANNs, the operation of convolution. Fast oscillating synchrony, may be the result of lateral connections of functionally adjacent neurons. Since sensory cortices are organized such that adjacent real world features are mapped spatially highly correlated to the cortex (see "Structural aspects of the brain"), almost simultaneously active and inter-connected neurons (that are allowed to synchronize by e.g. α dis-inhibition) lock into respective phases of γ . It has been observed in mice that a reduced number of inter-neurons lead to a decrease in δ , θ , α and γ power [460]. Hence, local connections of inter-neurons appear vital for synchronized cortical activity. Indeed, γ band oscillations have been linked to spike time coordination in the olfactory bulb [461]. Furthermore, it has been shown that γ band oscillation specifically couple functionally related neuronal assemblies [462]. Again, this coupling could be interpreted as the active processing within a neuronal assembly that has been activated and functionally coupled by feature-scale low frequency oscillations (α) and is further modified by object and feature scale low frequency oscillations (θ and α respectively). Additional evidence for the δ frequency range implies a role of δ in *temporal* sampling to ensure that objects are embedded into a temporal context [162]. Coincidentally, multiple θ cycles would fit into a single δ cycle which - if we extrapolate the principle - speaks in favor of an even more increased (temporal) kernel size of δ .

Figure 37 summarizes this principle in one schematic. Convolution is interpreted as the weighted sum of a subset of the input that is successively performed in order to transform the input, such that certain features are e.g. enhanced (similar to edge detection using Sobel filters [456]). Filter kernels are learned weights specifically responding to a certain type of input data arrangement. This is similar to what has been thought to be implemented via complex cells [30]. The combined information of a specific set of simple cells would be received by feature specific complex cells ⁴⁰. Here, complex cells are seen as dynamic entities that recruit specific low level neuronal populations that will predictably transform the input, such that some higher order value function is maximized. A prediction of the shape of the letter "A" for instance would cause a pre-stimulus recruitment of neuronal populations at the level of V1 coding for "/", "\", and "-" via feedback deep layer α coherence. Signals from neurons in α coherence are now set up to be transferred to the higher order region [91]. Since neuronal populations are not only synchronized vertically but furthermore horizontally, lateral in phase communication could happen as well, depending on the strength of the input signal and the correlation of the input signal. However, since locally adjacent input neurons are driven in a correlated manner anyway, due to adjacent feature to adjacent cortical area mapping (see "Structural aspects of the brain"), the further joint upwards modulation by decreased inhibition [91, 172, 224] finally causes some neurons to release an action potential. Due to lateral connectivity and the already correlated input strength of neighboring neurons, a burst of joint activity is released, aligning spike times of neurons, which causes γ band activity [461]. Since, the input signal arrives at L4 and deep layer feedback α is routed there [80, 82], incoming γ band activity (from previously synchronized feature processing populations) is nested into deep α , which transfers the information into the communication channel. In theory, this information could be routed via L2/3 of the sending region to L4 of the receiving region, however superficial attention α and potentially θ activity modulate the outgoing signal further. Thereby, attention α acts as a gate keeper and θ increases capacity of the visual processing by combining features into objects along its phase code [91]. A very much simplified representation, transferred to the working principle of a simple *Perceptron* [73] is proposed below. Note, that the lower order activation (γ_{low}) also depends on α and θ , even though the model does not directly reflect it. However, non-zero lower order activation (γ_{low}) would already be meaningfully influenced by low frequencies at a previous time step and is hence implicit to the model. For the simplified model below, the bias for each neuron is different, however biases of spatially and functionally adjacent neurons are correlated.

⁴⁰Whether they would be indeed cellular entities or functionally coupled cell assemblies to fulfill a functionally similar purpose is insignificant for the theory.

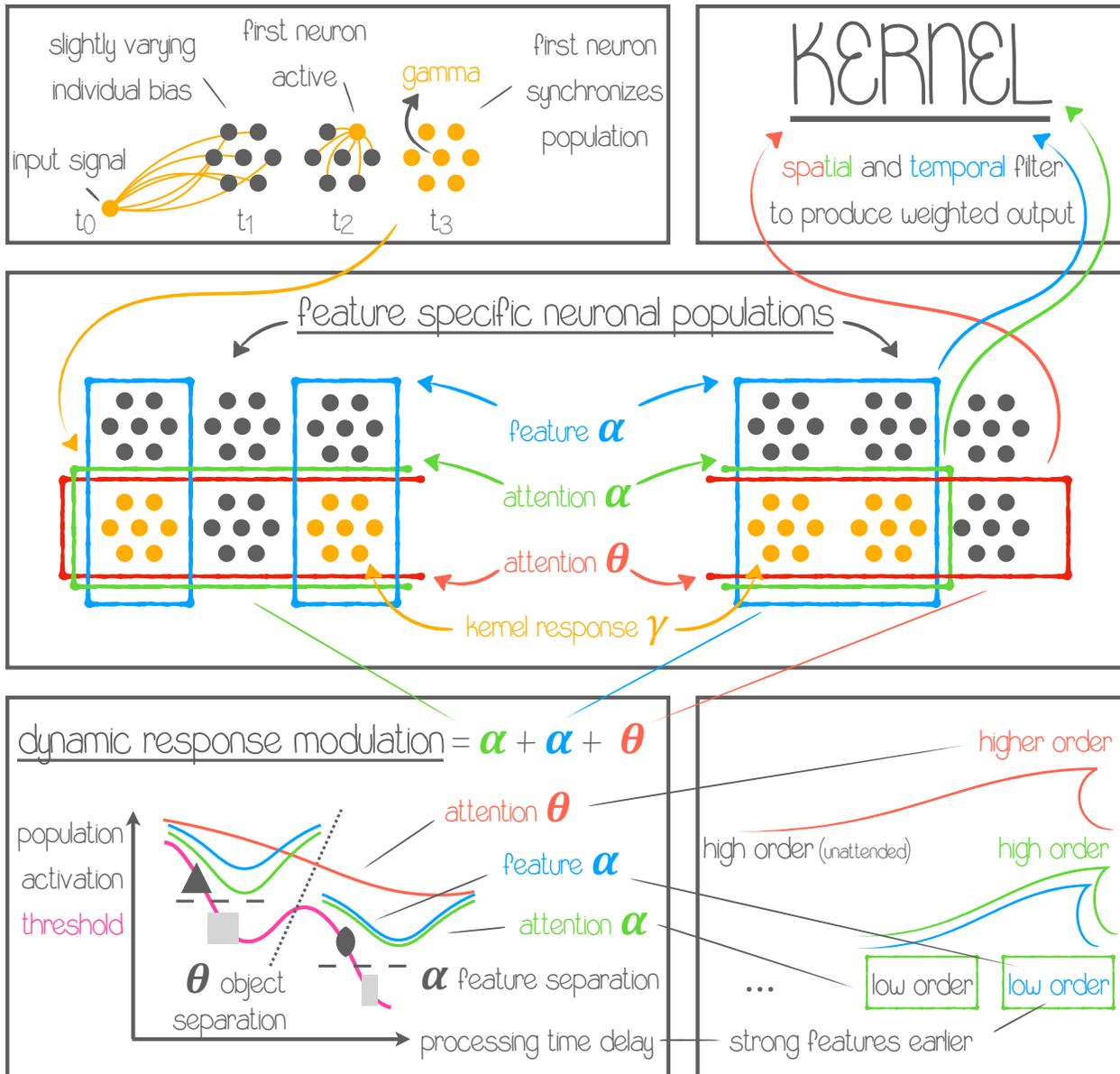


Figure 37: **The convolutional brain** For detailed explanation see page 176. Deep layer α recruits multiple feature specific populations of neurons, by anticipating the stimulus. Coherent α "samples" populations of neurons corresponding to the strength of the input signal. Top down superficial layer α and θ modulate the respective neuron selection such that a compound phase gradient groups responses to features and features to objects. Feature and attention α and θ thereby modulate the input signal such that informative features are enhanced (e.g. edges using a Sobel filter [456]). The process is thought similar to that of applying convolution filters in CNNs, but the respective filter width is determined by the opportunity window provided by the underlying frequency. Applying the filter necessarily means correlating the bias of spatially adjacent neurons and hence a few neurons hitting the threshold would (almost) instantaneously synchronize the entire population. This operation is hypothesized to be macroscopically visible as γ band activity.

$$f_{high}(t, \theta, \alpha_{att}, \alpha_{feat}, \gamma) = \begin{cases} 1 & \text{if } w_{low \rightarrow high} \cdot \gamma_{low}(t) + b(t) > 0, \\ 0 & \text{otherwise} \end{cases}$$

, with

$$b(t) = -(|\alpha_{feat}(t)| + |\alpha_{att}(t)| + |\theta(t)|)$$

, such that feature specific α (α_{feat}) recruits a set of neurons that transform the input into a filtered representation enhancing certain aspects of the stimulus. Thereby, the input signal (transformed into γ band oscillations (L4) via local synchronization) is nested into α_{feat} (L5/6) allowing lateral communication between simultaneously dis-inhibited neurons. Attention α (α_{att}) acts as a gatekeeper, modulating the strength of the respective γ transformed and α_{feat} nested output signal in L2/3. Long range θ connectivity binds or separates respective α_{feat} and α_{att} features into sets (objects). The lower right part of Figure 37 illustrates this process in a graph. Thereby, the x-axis represents the *delay* time at which a certain stimulus is processed. Phase coding frameworks propose that a α or θ phase gradient could be used to sort incoming signals according to their strength [91, 204]. Hence, strong stimuli would be processed *earlier* than weaker stimuli. Stimulus activation (strength of neuronal input) is reflected by the y-axis of the graph. A strictly linear relationship between input strength and activation delay would result in a line from top left to bottom right of the graph, which would reflect the activation threshold. Crossing the line towards the top right would thereby indicate neuronal signal transmission (i.e. action potential). Underlying α and θ oscillations however, modify the activation threshold of the neuronal populations rhythmically, such that the linear relationship is deflected according to the phase and amplitude relationship of α and θ . This convoluted processing has two major advantages: Filter response sorting along the α phase (if the input exposes features to which the selected neurons strongly respond, those features are processed earlier and the respective kernel becomes more important) and feature grouping (i.e. object separation) or feature group sampling (strongest average response of a feature group gets processed earlier in the θ phase). In a classical CNN, the kernel is moved over a respective domain spatially (with respect to the input neurons). Here, the kernel is thought to reflect a temporal "movement". Since the "kernel" would actually be a subset of neurons selected by feedback α coherence that is modulated by attention α and θ , shifting the kernel is done in the temporal domain. Sorted along the phase gradient of α , strongest active neurons (coding specific features, e.g. bar orientations; also known as simple cells [30]) are processed first, which - in the phase coding framework - sorts them according to their respective importance. Feature, specific "simple cell" responses (pre-modulated by α) synchronize local activation of adjacent neurons that are just below threshold (upper left part of Figure 37), nesting incoming spikes via γ into α . Hence, jointly active features (within one α dis-inhibition cycle) form a new higher order feature once transferred to higher order regions. This process is similar to what

complex cells have been thought to implement. Superficial attention driven long range top down feedback α and θ modulate the signal further, such that either important features get boosted or distractors depressed [251]. Attention related α is hypothesized to be suitable to implement multiple forms of attention, such as spatial, color, etc, depending on task demand and stimulus material. Hence, deep layer and superficial layer α might actually implement a similar functionality, where deep layer α reflects processing feedback for closely related processes (e.g. complex cell functionality) or more abstract top down influence (e.g. attention). Both aspects however modulate the signal (i.e. activation threshold) and thereby shape the "kernel". Transferred to the logic of classic CNNs, θ reflects the boosted activity of some top down attention, focusing on the e.g. foreground object. Thereby, θ would spatially bind neurons encoding the respective objects along its phase. In turn feature specific α would implement the major part of the convolution operation, by recruiting phase coherent neurons to implement a certain filter to the input. Sliding the filter across the input would be implemented by a varying response strength of filtered input signals, which are coded along α_{feat} . Lastly, α_{att} applies an importance weighting to the input, depending on the respective underlying value function.

Predictions In line with previous frameworks [91, 258] a first prediction that could directly be answered by completing the analysis of the EEG-fMRI experiment with respect to connectivity, is that, if indeed deep layer α recruits neurons feature specifically, such that the most informative features get enhanced relative to non informative features, then increased connectivity from top down deep layer regions to deep layer lower region should be observed, which is correlated with α band activity. Verification that such an approach might yield meaningful results has been provided recently [265]. Ideally, α band connectivity could be measured directly. This would be possible using laminar level MEG, which as well has been demonstrated to provide suitable spatial resolution [220, 268]. Since MEG data from both MEG experiments have been recorded with the respective spatial precision in mind, results on laminar connectivity can shed light on this question. It would be expected that deep layer α band connectivity specifically targets (functionally closely related) lower order regions. The face-name Stroop task (see "Overview Experiment 2: feature attention and conflicting information") has been set up to differentiate multiple higher order visual processing areas: FFA and VWFA, which draw on information from similar areas (e.g. V1). Hence, depending on the respective attention condition (attend-face vs attend-name) only neuronal populations - set up via deep layer α - that a) code task relevant feature information and b) are attended and therefore communicate coherently via a channel are selected. Respective α coherence between FFA and V1 or VWFA and V1 is thus expected to vary depending on the attention condition.

Since it is expected that γ band oscillations are nested into feedback α [91] and the present model predicts this to happen between L5/6 and L4, γ oscillations are expected to be related to the (at least) deep layer α phase, which could be tested via a corresponding analysis. In addition to that the interaction between feature selective α and γ is hypothesized to be modulated by attention α (and possibly θ) which could be tested based on localizer regions obtained from the fMRI data of respective experiments.

Furthermore, attention α and θ are thought to be reflected in long range connectivity [420, 457]. While both could be reflected in deep *and* superficial layers [52], findings from the present EEG-fMRI experiment (see "Feature specific neuronal oscillations in cortical layers (in prep)") suggest an influence of attention (or at least more general non-specific activation modulation) mainly in superficial layers. Whether or not this extrapolates to θ as well remains unclear, due to conflicting findings in the literature [297], but for the simplicity of the model it has been asserted to superficial layers as well, since a α / θ attention sampling mechanism is hypothesized anyway [210, 309, 457]. This means however that functional connectivity in the α *and* θ range predominantly between superficial layers of high order regions (e.g. FEF or PFC) is expected. Again the MEG data recorded using the face-name Stroop task could provide insight into this assumption. Theoretically, it should even be possible to decode attend-face and attend-name stimuli from long range α / θ connectivity, since the changes to the "kernel" would be object specific (i.e. for *word* or *face*). It has been shown that θ band activity in the visual cortex spreads predominantly in feed-forward direction [242] however a feedback evoked (from higher order regions e.g. FEF [251]) feed-forward sweep in primary visual regions [250] would likely be a consequence of synchronizing neuronal populations anyway.

Since γ oscillations are thought to reflect the "kernel response" (signal after spatio-temporal filtering), it is *not* expected to vary as a response to the *type* of kernel used (e.g. spatially repetitive; low order features, such as lines vs high order features, such as corners) - i.e. γ is not a kernel specific signature - but as a function of the stimulus to the kernel. Simple features - such as gratings - hence elicit large γ band responses (see e.g. "Feature specific neuronal oscillations in cortical layers (in prep)"), because the response to e.g. edge enhancing kernels is very strong (the stimulus is composed of edges only). It has been found that gratings with a higher spatial frequency, elicit stronger γ band responses than gratings with a lower spatial frequency [189]. Additionally, it has been found by the same study that complex stimuli result in a more broadband response, whereas simple stimuli (e.g. gratings) result in a more narrow band response. In the light of the proposed model, this would be the consequence of "kernel resonance". In other words, the match between neuronal populations coding for a set of features is stronger, the stronger the correspondence between the features and the kernel. It has been found that γ band synchronizations between IFJ and either FFA or PPA vary based on the whether faces or houses were presented [245]. For the face-name

Stroop MEG experiment, similar connectivity changes between primary visual regions and either FFA or VWFA (depending on whether a face or word has been attended) would be expected. Thereby, similarly strong γ band activity would be expected in a region where the stimulus presented is congruent to the specialization of the target area. Filter kernels are expected to anatomically (and thereby somewhat functionally) be the result of some sort of self organized learning, comparable to Kohonen maps [37] during development. Hence, filter kernels probably respond stronger to stimuli that occur often, as compared to stimuli that occur less often. The more a kernel proved that the filtered result is informative, the more distinct this kernel probably finds representation in the cortex. This means that familiar (i.e. common) objects should elicit stronger γ band responses than unfamiliar objects. Composing stimuli such that a respective regional increase in the narrow band γ can be observed, which could hence be a way to determine the shape of a respective filter kernel. One possibility would be to derive feature maps from CNNs, visualize the preferred kernel response and use those images as stimuli for future experiments. It would be predicted that at least some filters that a visual discrimination network has learned result in high narrow band γ band responses. Of course, those correspondences are nowhere near proving an actual correspondence, but could at least serve as a starting point. A common strategy to visualize filter weights in neural networks is to generate input images from traced back back-propagation errors that maximize the response [437]. If γ band oscillations indeed reflect the response of a set of neurons to certain input filter by feature specific neurons, then a potential way to visualize filter weights could be a neuro-feedback online BCI system, where a neural network is trained unsupervised to modify the input to a human subject inside the MEG, such that γ band activity in a specific region is maximized. In any case, γ band activity is expected to be coherent between low and higher order regions, nested in the α band [91].

Final remarks It has come to the authors attention that the proposed model might appear logical and plausible but turns out to be extremely difficult to test down to its core predictions (e.g. "real" filter kernels). Nevertheless, I believe that this viewing angle helps to view the brain more from a computational perspective that has implemented certain functions adjusted to environmental requirements. Thus, the presented framework shares high similarities (in form of predictions) with other contemporary frameworks [91, 240, 247, 379]. The major advantage of the proposed framework however is to provide at least a working hypothesis about the *why* of some functionally relevant neuronal (oscillatory) activity on the level of comprehensive neuronal network understanding. Viewing the brain as a CNN (at least on the level of primary sensory processing) provides a whole new set of testable hypothesis. Of course the framework in its current state can be seen as highly speculative. However, a holistic interpretation of the data almost necessarily comprises the influence of individual biases. Mine is biased towards the brain as a neural network.

Bibliography

- [1] Maryam Mostafalu. *Tracking the attentional spotlight with high precision MEG (H/F)*. PhD thesis, University Claude Bernard Lyon 1, in progress.
- [2] David G Robinson and Andreas Draguhn. Plants have neither synapses nor a nervous system. *Journal of Plant Physiology*, 263:153467, 2021.
- [3] Hridya Hemachandran, C George Priya Doss, and Ramamoorthy Siva. Plant communication: an unresolved mystery. *Current Science*, 112(10):1990, 2017.
- [4] Pablo H Maseda and Roberto J Fernández. Stay wet or else: three ways in which plants can adjust hydraulically to their environment. *Journal of experimental botany*, 57(15):3963–3977, 2006.
- [5] Muthusubramanian Venkateshwaran, Jeremy D Volkening, Michael R Sussman, and Jean-Michel Ané. Symbiosis and the social network of higher plants. *Current Opinion in Plant Biology*, 16(1):118–127, 2013.
- [6] Hiroshi Watanabe, Toshitaka Fujisawa, and Thomas W Holstein. Cnidarians and the evolutionary origin of the nervous system. *Development, growth & differentiation*, 51(3):167–183, 2009.
- [7] Deon Furstenburg and W Van Hoven. Condensed tannin as anti-defoliate agent against browsing by giraffe (*giraffa camelopardalis*) in the kruger national park. *Comparative Biochemistry and Physiology Part A: Physiology*, 107(2):425–431, 1994.
- [8] Timothy O’Connor. Emergent properties. *American Philosophical Quarterly*, 31(2):91–104, 1994.
- [9] Upinder S Bhalla and Ravi Iyengar. Emergent properties of networks of biological signaling pathways. *Science*, 283(5400):381–387, 1999.
- [10] Michael J Bok, Megan L Porter, Allen R Place, and Thomas W Cronin. Biological sunscreens tune polychromatic ultraviolet vision in mantis shrimp. *Current Biology*, 24(14):1636–1642, 2014.

- [11] Margaret Wilson. Six views of embodied cognition. *Psychonomic bulletin & review*, 9(4): 625–636, 2002.
- [12] Chad J Donahue, Matthew F Glasser, Todd M Preuss, James K Rilling, and David C Van Essen. Quantitative assessment of prefrontal cortex in humans relative to nonhuman primates. *Proceedings of the National Academy of Sciences*, 115(22):E5183–E5192, 2018.
- [13] Naomi P Friedman and Trevor W Robbins. The role of prefrontal cortex in cognitive control and executive function. *Neuropsychopharmacology*, 47(1):72–89, 2022.
- [14] Suzana Herculano-Houzel. The human brain in numbers: a linearly scaled-up primate brain. *Frontiers in human neuroscience*, page 31, 2009.
- [15] Sarah DeWeerd. How to map the brain. *Nature*, 571(7766):S6–S6, 2019.
- [16] Suzana Herculano-Houzel. The remarkable, yet not extraordinary, human brain as a scaled-up primate brain and its associated cost. *Proceedings of the National Academy of Sciences*, 109 (supplement_1):10661–10668, 2012.
- [17] Henry Markram. The human brain project. *Scientific American*, 306(6):50–55, 2012.
- [18] Clint Witchalls. A computer that thinks. *New Scientist*, 224(2994):28–29, 2014.
- [19] Peter Bain, Tipu Aziz, Xuguang Liu, and Dipankar Nandi. *Deep brain stimulation*. Oxford University Press, 2009.
- [20] N Alteheld, G Roessler, M Vobig, and P Walter. The retina implant new approach to a visual prosthesis/das retina-implant ein neuer ansatz für eine sehprothese. 2004.
- [21] Fan-Gang Zeng, Stephen Rebscher, William Harrison, Xiaoan Sun, and Haihong Feng. Cochlear implants: system design, integration, and evaluation. *IEEE reviews in biomedical engineering*, 1:115–142, 2008.
- [22] Andrew B Schwartz, X Tracy Cui, Douglas J Weber, and Daniel W Moran. Brain-controlled interfaces: movement restoration with neural prosthetics. *Neuron*, 52(1):205–220, 2006.
- [23] Charles N David. A quantitative method for maceration of hydra tissue. *Wilhelm Roux’Archiv für Entwicklungsmechanik der Organismen*, 171(4):259–268, 1973.
- [24] Christophe Dupre and Rafael Yuste. Non-overlapping neural networks in hydra vulgaris. *Current Biology*, 27(8):1085–1097, 2017.
- [25] Ni Ji and Steven W Flavell. Hydra: imaging nerve nets in action. *Current Biology*, 27(8): R294–R295, 2017.
- [26] Hideo Hasegawa. Dynamical mean-field theory of noisy spiking neuron ensembles: Application to the hodgkin-huxley model. *Physical Review E*, 68(4):041909, 2003.

- [27] Ernest Montbrió, Diego Pazó, and Alex Roxin. Macroscopic description for networks of spiking neurons. *Physical Review X*, 5(2):021028, 2015.
- [28] Halgurd Taher, Alessandro Torcini, and Simona Olmi. Exact neural mass model for synaptic-based working memory. *PLoS Computational Biology*, 16(12):e1008533, 2020.
- [29] Peter R Huttenlocher et al. Synaptic density in human frontal cortex-developmental changes and effects of aging. *Brain Res*, 163(2):195–205, 1979.
- [30] David H Hubel and Torsten N Wiesel. Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex. *The Journal of physiology*, 160(1):106, 1962.
- [31] David Marr and Ellen Hildreth. Theory of edge detection. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 207(1167):187–217, 1980.
- [32] Serge O. Dumoulin and Brian A. Wandell. Population receptive field estimates in human visual cortex. *NeuroImage*, 39(2):647–660, January 2008. ISSN 10538119. doi: 10.1016/j.neuroimage.2007.09.034. URL <http://linkinghub.elsevier.com/retrieve/pii/S1053811907008269>.
- [33] Mario Senden, Thomas C Emmerling, Rick Van Hoof, Martin A Frost, and Rainer Goebel. Reconstructing imagined letters from early visual cortex reveals tight topographic correspondence between visual mental imagery and perception. *Brain Structure and Function*, 224(3):1167–1183, 2019.
- [34] Peter Kok, Lauren J. Bains, Tim van Mourik, David G. Norris, and Floris P. de Lange. Selective Activation of the Deep Layers of the Human Primary Visual Cortex by Top-Down Feedback. *Current Biology*, 26(3):371–376, February 2016. ISSN 09609822. doi: 10.1016/j.cub.2015.12.038. URL <http://linkinghub.elsevier.com/retrieve/pii/S0960982215015699>.
- [35] William H Bosking, Ying Zhang, Brett Schofield, and David Fitzpatrick. Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex. *Journal of neuroscience*, 17(6):2112–2127, 1997.
- [36] Jean-Luc R Stevens, Judith S Law, Ján Antolík, and James A Bednar. Mechanisms for stable, robust, and adaptive development of orientation maps in the primary visual cortex. *Journal of Neuroscience*, 33(40):15747–15766, 2013.
- [37] Teuvo Kohonen. Self-organized formation of topologically correct feature maps. *Biological cybernetics*, 43(1):59–69, 1982.
- [38] Sanne Schoenmakers, Markus Barth, Tom Heskes, and Marcel Van Gerven. Linear reconstruction of perceived images from human brain activity. *NeuroImage*, 83:951–961, 2013.
- [39] Sanne Schoenmakers, Umut Güçlü, Marcel Van Gerven, and Tom Heskes. Gaussian mixture models and semantic gating improve reconstructions from human brain activity. *Frontiers in computational neuroscience*, 8:173, 2015.

- [40] Katja Seeliger, Umut Güçlü, Luca Ambrogioni, Yagmur Güçlütürk, and Marcel AJ van Gerven. Generative adversarial networks for reconstructing natural images from brain activity. *NeuroImage*, 181:775–785, 2018.
- [41] Wilder Penfield and Edwin Boldrey. Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. *Brain*, 60(4):389–443, 1937.
- [42] Elia Formisano, Dae-Shik Kim, Francesco Di Salle, Pierre-Francois Van de Moortele, Kamil Ugurbil, and Rainer Goebel. Mirror-symmetric tonotopic maps in human primary auditory cortex. *Neuron*, 40(4):859–869, 2003.
- [43] Melissa Saenz and Dave RM Langers. Tonotopic mapping of human auditory cortex. *Hearing research*, 307:42–52, 2014.
- [44] Kensaku Mori and Hitoshi Sakano. How is the olfactory map formed and interpreted in the mammalian brain? *Annual review of neuroscience*, 34:467–499, 2011.
- [45] Seiji Ogawa, Tso-Ming Lee, Alan R Kay, and David W Tank. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *proceedings of the National Academy of Sciences*, 87(24):9868–9872, 1990.
- [46] Steven B Most, Daniel J Simons, Brian J Scholl, Rachel Jimenez, Erin Clifford, and Christopher F Chabris. How not to be seen: The contribution of similarity and selective ignoring to sustained inattentive blindness. *Psychological science*, 12(1):9–17, 2001.
- [47] Sunil P Gandhi, David J Heeger, and Geoffrey M Boynton. Spatial attention affects brain activity in human primary visual cortex. *Proceedings of the National Academy of Sciences*, 96(6):3314–3319, 1999.
- [48] John J Foxe and Adam C Snyder. The role of alpha-band brain oscillations as a sensory suppression mechanism during selective attention. *Frontiers in psychology*, 2:154, 2011.
- [49] Mika H Martikainen, Ken-ichi Kaneko, and Riitta Hari. Suppressed responses to self-triggered sounds in the human auditory cortex. *Cerebral cortex*, 15(3):299–302, 2005.
- [50] Olaf Sporns and Richard F Betzel. Modular brain networks. *Annual review of psychology*, 67:613, 2016.
- [51] Victor AF Lamme, Hans Super, and Henk Spekreijse. Feedforward, horizontal, and feedback processing in the visual cortex. *Current opinion in neurobiology*, 8(4):529–535, 1998.
- [52] Timo van Kerkoerle, Matthew W. Self, Bruno Dagnino, Marie-Alice Gariel-Mathis, Jasper Poort, Chris van der Togt, and Pieter R. Roelfsema. Alpha and Gamma Oscillations Characterize Feedback and Feedforward Processing in Monkey Visual Cortex. *Proceedings of the National Academy of Sciences*, 111(40):14332–14341, October 2014. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1402773111.

- [53] Elizabeth A Buffalo, Pascal Fries, Rogier Landman, Timothy J Buschman, and Robert Desimone. Laminar differences in gamma and alpha coherence in the ventral stream. *Proceedings of the National Academy of Sciences*, 108(27):11262–11267, 2011.
- [54] Troy A Hackett. Information flow in the auditory cortical network. *Hearing research*, 271(1-2):133–146, 2011.
- [55] Rachael D Seidler, Douglas C Noll, and G Thiers. Feedforward and feedback processes in motor control. *Neuroimage*, 22(4):1775–1783, 2004.
- [56] Sabine Kastner, Mark A Pinsk, Peter De Weerd, Robert Desimone, and Leslie G Ungerleider. Increased activity in human visual cortex during directed attention in the absence of visual stimulation. *Neuron*, 22(4):751–761, 1999.
- [57] Jorge F Mejias, John D Murray, Henry Kennedy, and Xiao-Jing Wang. Feedforward and feedback frequency-dependent interactions in a large-scale laminar network of the primate cortex. *Science advances*, 2(11):e1601335, 2016.
- [58] Jonathan Smallwood, Boris C Bernhardt, Robert Leech, Danilo Bzdok, Elizabeth Jefferies, and Daniel S Margulies. The default mode network in cognition: a topographical perspective. *Nature reviews neuroscience*, 22(8):503–513, 2021.
- [59] Kaspar Meyer and Antonio Damasio. Convergence and divergence in a neural architecture for recognition and memory. *Trends in neurosciences*, 32(7):376–382, 2009.
- [60] Kingson Man, Jonas Kaplan, Hanna Damasio, and Antonio Damasio. Neural convergence and divergence in the mammalian cerebral cortex: from experimental neuroanatomy to functional neuroimaging. *Journal of Comparative Neurology*, 521(18):4097–4111, 2013.
- [61] Jean Bullier and Lionel G Nowak. Parallel versus serial processing: new vistas on the distributed organization of the visual system. *Current opinion in neurobiology*, 5(4):497–503, 1995.
- [62] Ludovic S Mure. Intrinsically photosensitive retinal ganglion cells of the human retina. *Frontiers in neurology*, 12:636330, 2021.
- [63] HeB Barlow, RM Hill, and WR Levick. Retinal ganglion cells responding selectively to direction and speed of image motion in the rabbit. *The Journal of physiology*, 173(3):377, 1964.
- [64] Shijun Weng, Wenzhi Sun, and Shigang He. Identification of on–off direction-selective ganglion cells in the mouse retina. *The Journal of physiology*, 562(3):915–923, 2005.
- [65] Dennis M Dacey. Primate retina: cell types, circuits and color opponency. *Progress in retinal and eye research*, 18(6):737–763, 1999.

- [66] EJ Chichilnisky and Rachel S Kalmar. Functional asymmetries in on and off ganglion cells of primate retina. *Journal of Neuroscience*, 22(7):2737–2747, 2002.
- [67] Takao Hashimoto, Satoshi Katai, Yasunori Saito, Fumitoshi Kobayashi, and Tetsuya Goto. On and off channels in human retinal ganglion cells. *The Journal of Physiology*, 591(1):327–337, 2013.
- [68] W Martin Usrey and Henry J Alitto. Visual functions of the thalamus. *Annual review of vision science*, 1:351, 2015.
- [69] Adam M Sillito, Javier Cudeiro, and Helen E Jones. Always returning: feedback and sensory processing in visual cortex and thalamus. *Trends in neurosciences*, 29(6):307–316, 2006.
- [70] Farran Briggs and W Martin Usrey. Corticogeniculate feedback and visual processing in the primate. *The Journal of physiology*, 589(1):33–40, 2011.
- [71] Wayne E Mackey, Jonathan Winawer, and Clayton E Curtis. Visual field map clusters in human frontoparietal cortex. *Elife*, 6:e22974, 2017.
- [72] Brian A Wandell, Serge O Dumoulin, and Alyssa A Brewer. Visual field maps in human cortex. *Neuron*, 56(2):366–383, 2007.
- [73] Frank Rosenblatt. *The perceptron, a perceiving and recognizing automaton Project Para*. Cornell Aeronautical Laboratory, 1957.
- [74] Nikolaus Kriegeskorte. Deep neural networks: a new framework for modelling biological vision and brain information processing. *bioRxiv*, page 029876, 2015.
- [75] Carolyn Jeane Perry and Mazyar Fallah. Feature integration and object representations along the dorsal stream visual hierarchy. *Frontiers in computational neuroscience*, 8:84, 2014.
- [76] Tirin Moore and Katherine M Armstrong. Selective gating of visual signals by microstimulation of frontal cortex. *Nature*, 421(6921):370–373, 2003.
- [77] Brad C Motter. Focal attention produces spatially selective processing in visual cortical areas v1, v2, and v4 in the presence of competing stimuli. *Journal of neurophysiology*, 70(3):909–919, 1993.
- [78] Simon Clavagnier, Arnaud Falchier, and Henry Kennedy. Long-distance feedback projections to area v1: implications for multisensory integration, spatial awareness, and visual consciousness. *Cognitive, Affective, & Behavioral Neuroscience*, 4(2):117–126, 2004.
- [79] S Shushruth, Pradeep Mangapathy, Jennifer M Ichida, Paul C Bressloff, Lars Schwabe, and Alessandra Angelucci. Strong recurrent networks compute the orientation tuning of surround modulation in the primate primary visual cortex. *Journal of Neuroscience*, 32(1):308–321, 2012.

- [80] Nikola T. Markov, Julien Vezoli, Pascal Chameau, Arnaud Falchier, René Quilodran, Cyril Huisoud, Camille Lamy, Pierre Misery, Pascale Giroud, Shimon Ullman, Pascal Barone, Colette Dehay, Kenneth Knoblauch, and Henry Kennedy. Anatomy of hierarchy: Feedforward and feedback pathways in macaque visual cortex: Cortical counterstreams. *Journal of Comparative Neurology*, 522(1):225–259, January 2014. ISSN 00219967. doi: 10.1002/cne.23458. URL <http://doi.wiley.com/10.1002/cne.23458>.
- [81] Pascal Barone, Alexandre Batardiere, Kenneth Knoblauch, and Henry Kennedy. Laminar distribution of neurons in extrastriate areas projecting to visual areas v1 and v4 correlates with the hierarchical rank and indicates the operation of a distance rule. *Journal of Neuroscience*, 20(9):3263–3281, 2000.
- [82] Andre M Bastos, W Martin Usrey, Rick A Adams, George R Mangun, Pascal Fries, and Karl J Friston. Canonical microcircuits for predictive coding. *Neuron*, 76(4):695–711, 2012.
- [83] W Martin Usrey and David Fitzpatrick. Specificity in the axonal connections of layer vi neurons in tree shrew striate cortex: evidence for distinct granular and supragranular systems. *Journal of Neuroscience*, 16(3):1203–1218, 1996.
- [84] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *nature*, 323(6088):533–536, 1986.
- [85] João Sacramento, Rui Ponte Costa, Yoshua Bengio, and Walter Senn. Dendritic cortical microcircuits approximate the backpropagation algorithm. *Advances in neural information processing systems*, 31, 2018.
- [86] Stephen Grossberg. Competitive learning: From interactive activation to adaptive resonance. *Cognitive science*, 11(1):23–63, 1987.
- [87] Francis Crick. The recent excitement about neural networks. *Nature*, 337(6203):129–132, 1989.
- [88] Matthew Larkum. A cellular mechanism for cortical associations: an organizing principle for the cerebral cortex. *Trends in neurosciences*, 36(3):141–151, 2013.
- [89] K Brodman. *Vergleichende Lokalisationslehre der Grosshirnrinde, in ihren Prinzipien dargestellt auf Grund des Zellenbaues*. Barth Leipzig, 1909.
- [90] Sarah F Beul and Claus C Hilgetag. Towards a “canonical” agranular cortical microcircuit. *Frontiers in neuroanatomy*, 8:165, 2015.
- [91] Mathilde Bonnefond, Sabine Kastner, and Ole Jensen. Communication between Brain Areas Based on Nested Oscillations. *eneuro*, 4(2):ENEURO.0153–16.2017, 2017. ISSN 2373-2822. doi: 10.1523/ENEURO.0153-16.2017. URL <http://eneuro.sfn.org/lookup/doi/10.1523/ENEURO.0153-16.2017>.

- [92] Fioravante Capone, Matteo Paolucci, Federica Assenza, Nicoletta Brunelli, Lorenzo Ricci, Lucia Florio, and Vincenzo Di Lazzaro. Canonical cortical circuits: current evidence and theoretical implications. *Neuroscience and Neuroeconomics*, 5:1–8, 2016.
- [93] Rodney J Douglas, Kevan AC Martin, and David Whitteridge. A canonical microcircuit for neocortex. *Neural computation*, 1(4):480–488, 1989.
- [94] David C Godlove, Alexander Maier, Geoffrey F Woodman, and Jeffrey D Schall. Microcircuitry of agranular frontal cortex: testing the generality of the canonical cortical microcircuit. *Journal of Neuroscience*, 34(15):5355–5369, 2014.
- [95] Tirin Moore and Mazyar Fallah. Microstimulation of the frontal eye field and its effects on covert spatial attention. *Journal of neurophysiology*, 91(1):152–162, 2004.
- [96] Jakob Heinzle, Klaus Hepp, and Kevan AC Martin. A microcircuit model of the frontal eye fields. *Journal of Neuroscience*, 27(35):9341–9353, 2007.
- [97] Jelle A van Dijk, Alessio Fracasso, Natalia Petridou, and Serge O Dumoulin. Laminar processing of numerosity supports a canonical cortical microcircuit in human parietal cortex. *Current Biology*, 31(20):4635–4640, 2021.
- [98] McCulloch JL. A logical calculus of ideas immanent in nervous activity. *Bull. of Math. Biophysics*, 5:115–133, 1943.
- [99] Donald O Hebb. The first stage of perception: growth of the assembly. *The Organization of Behavior*, 4:60–78, 1949.
- [100] Bruce P Bean. The action potential in mammalian central neurons. *Nature Reviews Neuroscience*, 8(6):451–465, 2007.
- [101] Charles J Wilson and Philip M Groves. Spontaneous firing patterns of identified spiny neurons in the rat neostriatum. *Brain research*, 220(1):67–80, 1981.
- [102] Larry F Abbott. Lapicque’s introduction of the integrate-and-fire model neuron (1907). *Brain research bulletin*, 50(5-6):303–304, 1999.
- [103] Alan L Hodgkin and Andrew F Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of physiology*, 117(4):500, 1952.
- [104] Eugene M Izhikevich. Which model to use for cortical spiking neurons? *IEEE transactions on neural networks*, 15(5):1063–1070, 2004.
- [105] Eugene M Izhikevich. Simple model of spiking neurons. *IEEE Transactions on neural networks*, 14(6):1569–1572, 2003.

- [106] Michael Beyeler, Micah Richert, Nikil D Dutt, and Jeffrey L Krichmar. Efficient spiking neural network model of pattern motion selectivity in visual cortex. *Neuroinformatics*, 12(3): 435–454, 2014.
- [107] QingXiang Wu, T Martin McGinnity, Liam Maguire, Rongtai Cai, and Meigui Chen. A visual attention model based on hierarchical spiking neural networks. *Neurocomputing*, 116: 3–12, 2013.
- [108] Kristofor D Carlson, Micah Richert, Nikil Dutt, and Jeffrey L Krichmar. Biologically plausible models of homeostasis and stdp: stability and learning in spiking neural networks. In *The 2013 international joint conference on neural networks (IJCNN)*, pages 1–8. IEEE, 2013.
- [109] Aboozar Taherkhani, Ammar Belatreche, Yuhua Li, Georgina Cosma, Liam P Maguire, and T Martin McGinnity. A review of learning in biologically plausible spiking neural networks. *Neural Networks*, 122:253–272, 2020.
- [110] Jibin Wu, Yansong Chua, and Haizhou Li. A biologically plausible speech recognition framework based on spiking neural networks. In *2018 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2018.
- [111] Kashu Yamazaki, Viet-Khoa Vo-Ho, Darshan Bulsara, and Ngan Le. Spiking neural networks and their applications: A review. *Brain Sciences*, 12(7):863, 2022.
- [112] Paula Sanz Leon, Stuart A Knock, M Marmaduke Woodman, Lia Domide, Jochen Mersmann, Anthony R McIntosh, and Viktor Jirsa. The virtual brain: a simulator of primate brain network dynamics. *Frontiers in neuroinformatics*, 7:10, 2013.
- [113] Léon Bottou. Online algorithms and stochastic approximations. *Online learning and neural networks*, 1998.
- [114] Yann LeCun, Bernhard Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne Hubbard, and Lawrence D Jackel. Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4):541–551, 1989.
- [115] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.
- [116] David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, et al. Mastering the game of go without human knowledge. *nature*, 550(7676):354–359, 2017.
- [117] Sean D Holcomb, William K Porter, Shaun V Ault, Guifen Mao, and Jin Wang. Overview on deepmind and its alphago zero ai. In *Proceedings of the 2018 international conference on big data and education*, pages 67–71, 2018.

- [118] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.
- [119] Yuanqing Lin, Fengjun Lv, Shenghuo Zhu, Ming Yang, Timothee Cour, Kai Yu, Liangliang Cao, and Thomas Huang. Large-scale image classification: fast feature extraction and svm training. In *CVPR 2011*, pages 1689–1696. IEEE, 2011.
- [120] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2017.
- [121] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.
- [122] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [123] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pages 1026–1034, 2015.
- [124] Katja Seeliger, Luca Ambrogioni, Yağmur Güçlütürk, Leonieke M van den Bulk, Umut Güçlü, and Marcel AJ van Gerven. End-to-end neural system identification with neural information flow. *PLOS Computational Biology*, 17(2):e1008558, 2021.
- [125] Ken Chatfield, Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Return of the devil in the details: Delving deep into convolutional nets. *arXiv preprint arXiv:1405.3531*, 2014.
- [126] Katja Seeliger, Matthias Fritsche, Umut Güçlü, Sanne Schoenmakers, J-M Schoffelen, Sander E Bosch, and MAJ Van Gerven. Convolutional neural network-based encoding and decoding of visual object recognition in space and time. *NeuroImage*, 180:253–266, 2018.
- [127] Michael Eickenberg, Alexandre Gramfort, Gaël Varoquaux, and Bertrand Thirion. Seeing it all: Convolutional network layers map the function of the human visual system. *NeuroImage*, 152:184–194, 2017.
- [128] Nicholas J Sexton and Bradley C Love. Reassessing hierarchical correspondences between brain and deep networks through direct interface. *Science advances*, 8(28):eabm2219, 2022.
- [129] Eric Elmoznino and Michael F Bonner. High-performing neural network models of visual cortex benefit from high latent dimensionality. *bioRxiv*, 2022.

- [130] Rylan Schaeffer, Mikail Khona, and Ila Fiete. No free lunch from deep learning in neuroscience: A case study through models of the entorhinal-hippocampal circuit. *bioRxiv*, 2022.
- [131] Yanping Huang and Rajesh PN Rao. Predictive coding. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2(5):580–593, 2011.
- [132] Bhavin Choksi, Milad Mozafari, Callum Biggs O’May, Benjamin Ador, Andrea Alamia, and Rufin VanRullen. Predify: Augmenting deep neural networks with brain-inspired predictive coding dynamics. *Advances in Neural Information Processing Systems*, 34:14069–14083, 2021.
- [133] Rufin VanRullen and Ryota Kanai. Deep learning and the global workspace theory. *Trends in Neurosciences*, 44(9):692–704, 2021.
- [134] Kunihiko Fukushima and Sei Miyake. Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition. In *Competition and cooperation in neural nets*, pages 267–285. Springer, 1982.
- [135] Grace W Lindsay. Convolutional neural networks as a model of the visual system: Past, present, and future. *Journal of cognitive neuroscience*, 33(10):2017–2031, 2021.
- [136] James R Stieger, Stephen A Engel, Daniel Suma, and Bin He. Benefits of deep learning classification of continuous noninvasive brain–computer interface control. *Journal of neural engineering*, 18(4):046082, 2021.
- [137] Nibras Abo Alzahab, Luca Apollonio, Angelo Di Iorio, Muaaz Alshalak, Sabrina Iarlori, Francesco Ferracuti, Andrea Monteriù, and Camillo Porcaro. Hybrid deep learning (hdl)-based brain-computer interface (bci) systems: a systematic review. *Brain sciences*, 11(1):75, 2021.
- [138] Man Li, Feng Li, Jiahui Pan, Dengyong Zhang, Suna Zhao, Jingcong Li, and Fei Wang. The mindgomoku: An online p300 bci game based on bayesian deep learning. *Sensors*, 21(5):1613, 2021.
- [139] Qiulin Wang. Research on the improved cnn deep learning method for motion intention recognition of dynamic lower limb prosthesis. *Journal of Healthcare Engineering*, 2021, 2021.
- [140] Ali H Al-Timemy, Claudio Castellini, Javier Escudero, Rami Khushaba, and Silvia Muceli. Current trends in deep learning for movement analysis and prosthesis control. *Frontiers in Neuroscience*, 16, 2022.
- [141] Hojin Jang, Sergey M Plis, Vince D Calhoun, and Jong-Hwan Lee. Task-specific feature extraction and classification of fmri volumes using a deep neural network initialized with a deep belief network: Evaluation using sensorimotor tasks. *NeuroImage*, 145:314–328, 2017.

- [142] Salman Ul Hassan Dar, Muzaffer Özbey, Ahmet Burak Çatlı, and Tolga Çukur. A transfer-learning approach for accelerated mri using deep neural networks. *Magnetic resonance in medicine*, 84(2):663–685, 2020.
- [143] Rafael Elul. The genesis of the eeg. *International review of neurobiology*, 15:227–272, 1972.
- [144] Firas Fahoum, Renaud Lopes, Francesca Pittau, François Dubeau, and Jean Gotman. Widespread epileptic networks in focal epilepsies: Eeg-fmri study. *Epilepsia*, 53(9):1618–1627, 2012.
- [145] Claude Elwood Shannon. A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423, 1948.
- [146] Gyorgy Buzsaki and Andreas Draguhn. Neuronal oscillations in cortical networks. *science*, 304(5679):1926–1929, 2004.
- [147] Alison Hanson. Spontaneous electrical low-frequency oscillations: a possible role in hydra and all living systems. *Philosophical Transactions of the Royal Society B*, 376(1820):20190763, 2021.
- [148] Johanna M Zumer, René Scheeringa, Jan-Mathijs Schoffelen, David G Norris, and Ole Jensen. Occipital alpha activity during stimulus processing gates the information flow to object-selective cortex. *PLoS biology*, 12(10):e1001965, 2014.
- [149] Miranda Scolarì, Katharina N Seidl-Rathkopf, and Sabine Kastner. Functions of the human frontoparietal attention network: Evidence from neuroimaging. *Current opinion in behavioral sciences*, 1:32–39, 2015.
- [150] KE Stephan, KJ Friston, and LR Squire. Functional connectivity. 2009.
- [151] Karl J Friston, Chris D Frith, Peter F Liddle, and Richard SJ Frackowiak. Functional connectivity: the principal-component analysis of large (pet) data sets. *Journal of Cerebral Blood Flow & Metabolism*, 13(1):5–14, 1993.
- [152] André M Bastos and Jan-Mathijs Schoffelen. A tutorial review of functional connectivity analysis methods and their interpretational pitfalls. *Frontiers in systems neuroscience*, 9:175, 2016.
- [153] Clive WJ Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, pages 424–438, 1969.
- [154] Alexander Kraskov, Harald Stögbauer, and Peter Grassberger. Estimating mutual information. *Physical review E*, 69(6):066138, 2004.
- [155] Karl J Friston, Lee Harrison, and Will Penny. Dynamic causal modelling. *Neuroimage*, 19(4):1273–1302, 2003.

- [156] Stefan J Kiebel, Marta I Garrido, Rosalyn J Moran, and Karl J Friston. Dynamic causal modelling for eeg and meg. *Cognitive neurodynamics*, 2(2):121–136, 2008.
- [157] Olivier David, Stefan J Kiebel, Lee M Harrison, Jérémie Mattout, James M Kilner, and Karl J Friston. Dynamic causal modeling of evoked responses in eeg and meg. *NeuroImage*, 30(4):1255–1272, 2006.
- [158] Vangelis Sakkalis. Review of advanced techniques for the estimation of brain connectivity measured with eeg/meg. *Computers in biology and medicine*, 41(12):1110–1117, 2011.
- [159] Tzvetan Popov and Paul Szyszka. Alpha oscillations govern interhemispheric spike timing coordination in the honey bee brain. *Proceedings of the Royal Society B*, 287(1921):20200115, 2020.
- [160] Nobuaki K Tanaka, Kei Ito, and Mark Stopfer. Odor-evoked neural oscillations in drosophila are mediated by widely branching interneurons. *Journal of Neuroscience*, 29(26):8595–8603, 2009.
- [161] Nicolas Brunel. Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. *Journal of computational neuroscience*, 8(3):183–208, 2000.
- [162] Benjamin Morillon, Luc H Arnal, Charles E Schroeder, and Anne Keitel. Prominence of delta oscillatory rhythms in the motor cortex and their relevance for auditory and speech perception. *Neuroscience & Biobehavioral Reviews*, 107:136–142, 2019.
- [163] Freek van Ede, Andrew J Quinn, Mark W Woolrich, and Anna C Nobre. Neural oscillations: sustained rhythms or transient burst-events? *Trends in neurosciences*, 41(7):415–417, 2018.
- [164] Stuart W Hughes, Magor Lörincz, David W Cope, Kate L Blethyn, Katalin A Kékesi, H Rheinallt Parri, Gábor Juhász, and Vincenzo Crunelli. Synchronized oscillations at α and θ frequencies in the lateral geniculate nucleus. *Neuron*, 42(2):253–268, 2004.
- [165] Hans Berger. Über das elektroencephalogramm des menschen. *Archiv für psychiatrie und nervenkrankheiten*, 87(1):527–570, 1929.
- [166] Gert Pfurtscheller, A Stancak Jr, and Ch Neuper. Event-related synchronization (ers) in the alpha band—an electrophysiological correlate of cortical idling: a review. *International journal of psychophysiology*, 24(1-2):39–46, 1996.
- [167] Simon Hanslmayr, Tobias Staudigl, and Marie-Christin Fellner. Oscillatory power decreases and long-term memory: the information via desynchronization hypothesis. *Frontiers in human neuroscience*, 6:74, 2012.
- [168] C Dockstader, D Cheyne, and R Tannock. Cortical dynamics of selective attention to somatosensory events. *Neuroimage*, 49(2):1777–1785, 2010.

- [169] Sepideh Sadaghiani and Andreas Kleinschmidt. Brain networks and α -oscillations: structural and functional foundations of cognitive control. *Trends in cognitive sciences*, 20(11):805–817, 2016.
- [170] Lisa Payne and Robert Sekuler. The importance of ignoring: Alpha oscillations protect selectivity. *Current directions in psychological science*, 23(3):171–177, 2014.
- [171] Lisa Payne, Sylvia Guillory, and Robert Sekuler. Attention-modulated alpha-band oscillations protect against intrusion of irrelevant information. *Journal of cognitive neuroscience*, 25(9):1463–1476, 2013.
- [172] Ole Jensen, Bart Gips, Til Ole Bergmann, and Mathilde Bonnefond. Temporal coding organized by coupled alpha and gamma oscillations prioritize visual processing. *Trends in neurosciences*, 37(7):357–369, 2014.
- [173] Barbara F Händel, Thomas Haarmeier, and Ole Jensen. Alpha oscillations correlate with the successful inhibition of unattended stimuli. *Journal of cognitive neuroscience*, 23(9):2494–2502, 2011.
- [174] Paul Sauseng, Wolfgang Klimesch, Kirstin F Heise, Walter R Gruber, Elisa Holz, Ahmed A Karim, Mark Glennon, Christian Gerloff, Niels Birbaumer, and Friedhelm C Hummel. Brain oscillatory substrates of visual short-term memory capacity. *Current biology*, 19(21):1846–1852, 2009.
- [175] Plamen A Antonov, Ramakrishna Chakravarthi, and Søren K Andersen. Too little, too late, and in the wrong place: Alpha band activity does not reflect an active mechanism of selective attention. *Neuroimage*, 219:117006, 2020.
- [176] Joshua J Foster and Edward Awh. The role of alpha oscillations in spatial attention: limited evidence for a suppression account. *Current Opinion in Psychology*, 29:34–40, October 2019. ISSN 2352250X. doi: 10.1016/j.copsyc.2018.11.001. URL <https://linkinghub.elsevier.com/retrieve/pii/S2352250X18301684>.
- [177] Dirk van Moorselaar, Nasim Daneshmand, and Heleen A Slagter. Neural mechanisms underlying distractor inhibition on the basis of feature and/or spatial expectations. *Cortex*, 137:232–250, 2021.
- [178] Michael S Clayton, Nick Yeung, and Roi Cohen Kadosh. Electrical stimulation of alpha oscillations stabilizes performance on visual attention tasks. *Journal of Experimental Psychology: General*, 148(2):203, 2019.
- [179] Giovanni Piantoni, Nico Romeijn, German Gomez-Herrero, Ysbrand D Van Der Werf, and Eus JW Van Someren. Alpha power predicts persistence of bistable perception. *Scientific reports*, 7(1):1–11, 2017.

- [180] Corentin Gaillard and Suliann Ben Hamed. The neural bases of spatial attention and perceptual rhythms. *European Journal of Neuroscience*, 55(11-12):3209–3223, 2022.
- [181] Poppy Sharp, Tjerk Gutteling, David Melcher, and Clayton Hickey. Spatial attention tunes temporal processing in early visual cortex by speeding and slowing alpha oscillations. *Journal of Neuroscience*, 2022.
- [182] Wolf Singer and Charles M Gray. Visual feature integration and the temporal correlation hypothesis. *Annual review of neuroscience*, 18(1):555–586, 1995.
- [183] Catherine Tallon-Baudry and Olivier Bertrand. Oscillatory gamma activity in humans and its role in object representation. *Trends in cognitive sciences*, 3(4):151–162, 1999.
- [184] Olivier Bertrand and Catherine Tallon-Baudry. Oscillatory gamma activity in humans: a possible role for object representation. *International Journal of Psychophysiology*, 38(3):211–223, 2000.
- [185] Pascal Fries, Danko Nikolić, and Wolf Singer. The gamma cycle. *Trends in neurosciences*, 30(7):309–316, 2007.
- [186] Pieter R Roelfsema, Victor AF Lamme, and Henk Spekreijse. Synchrony and covariation of firing rates in the primary visual cortex during contour grouping. *Nature neuroscience*, 7(9):982–991, 2004.
- [187] Pascal Fries, Jan-Hinrich Schröder, Pieter R Roelfsema, Wolf Singer, and Andreas K Engel. Oscillatory neuronal synchronization in primary visual cortex as a correlate of stimulus selection. *Journal of Neuroscience*, 22(9):3739–3754, 2002.
- [188] Supratim Ray and John HR Maunsell. Do gamma oscillations play a role in cerebral cortex? *Trends in cognitive sciences*, 19(2):78–85, 2015.
- [189] Dora Hermes, KJ Miller, BA Wandell, and Jonathan Winawer. Stimulus dependence of gamma oscillations in human visual cortex. *Cerebral cortex*, 25(9):2951–2959, 2015.
- [190] Eleonora Bartoli, William Bosking, Yvonne Chen, Ye Li, Sameer A Sheth, Michael S Beauchamp, Daniel Yoshor, and Brett L Foster. Functionally distinct gamma range activity revealed by stimulus tuning in human visual cortex. *Current Biology*, 29(20):3345–3358, 2019.
- [191] Marvin S Keshner. 1/f noise. *Proceedings of the IEEE*, 70(3):212–218, 1982.
- [192] Supratim Ray and John HR Maunsell. Different origins of gamma rhythm and high-gamma activity in macaque visual cortex. *PLoS biology*, 9(4):e1000610, 2011.
- [193] Aman B Saleem, Anthony D Lien, Michael Krumin, Bilal Haider, Miroslav Román Rosón, Asli Ayaz, Kimberly Reinhold, Laura Busse, Matteo Carandini, and Kenneth D Harris. Subcortical source and modulation of the narrowband gamma oscillation in mouse visual cortex. *Neuron*, 93(2):315–322, 2017.

- [194] György Buzsáki. Theta oscillations in the hippocampus. *Neuron*, 33(3):325–340, 2002.
- [195] H Petsche, Ch Stumpf, and G Gogolak. The significance of the rabbit’s septum as a relay station between the midbrain and the hippocampus i. the control of hippocampus arousal activity by the septum cells. *Electroencephalography and clinical neurophysiology*, 14(2):202–211, 1962.
- [196] John O’Keefe and Michael L Recce. Phase relationship between hippocampal place units and the eeg theta rhythm. *Hippocampus*, 3(3):317–330, 1993.
- [197] John O’keefe and Lynn Nadel. Précis of o’keefe & nadel’s the hippocampus as a cognitive map. *Behavioral and Brain Sciences*, 2(4):487–494, 1979.
- [198] John O’Keefe. Place units in the hippocampus of the freely moving rat. *Experimental neurology*, 51(1):78–109, 1976.
- [199] Rune W Berg and David Kleinfeld. Rhythmic whisking by rat: retraction as well as protraction of the vibrissae is under active muscular control. *Journal of neurophysiology*, 89(1):104–117, 2003.
- [200] Jorge Otero-Millan, Xoana G Troncoso, Stephen L Macknik, Ignacio Serrano-Pedraza, and Susana Martinez-Conde. Saccades and microsaccades during visual fixation, exploration, and search: foundations for a common saccadic generator. *Journal of vision*, 8(14):21–21, 2008.
- [201] Usha Goswami. A temporal sampling framework for developmental dyslexia. *Trends in cognitive sciences*, 15(1):3–10, 2011.
- [202] William E Skaggs, Bruce L McNaughton, Matthew A Wilson, and Carol A Barnes. Theta phase precession in hippocampal neuronal populations and the compression of temporal sequences. *Hippocampus*, 6(2):149–172, 1996.
- [203] W Klimesch and M Doppelmayr. encoding of new. *Neuroreport*, 7:1235–1240, 1996.
- [204] John E Lisman and Ole Jensen. The theta-gamma neural code. *Neuron*, 77(6):1002–1016, 2013.
- [205] Niko A Busch, Julien Dubois, and Rufin VanRullen. The phase of ongoing eeg oscillations predicts visual perception. *Journal of neuroscience*, 29(24):7869–7876, 2009.
- [206] Qiaoli Huang and Huan Luo. Saliency-based rhythmic coordination of perceptual predictions. *Journal of Cognitive Neuroscience*, 32(2):201–211, 2020.
- [207] Rufin VanRullen. Attention cycles. *Neuron*, 99(4):632–634, 2018.
- [208] Justin Riddle, Jason M Scimeca, Dillan Cellier, Sofia Dhanani, and Mark D’Esposito. Causal evidence for a role of theta and alpha oscillations in the control of working memory. *Current Biology*, 30(9):1748–1754, 2020.

- [209] PG Pavithran, K Arunkumar, NP Guhan Seshadri, Bikesh Kumar Singh, V Mahesh, and B Geethanjali. Index of theta/alpha ratio to quantify visual-spatial attention in dyslexics using electroencephalogram. In *2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS)*, pages 417–422. IEEE, 2019.
- [210] Ian C Fiebelkorn, Yuri B Saalman, and Sabine Kastner. Rhythmic sampling within and between objects despite sustained attention at a cued location. *Current biology*, 23(24):2553–2558, 2013.
- [211] Ravi V Chacko, Byungchan Kim, Suh Woo Jung, Amy L Daitch, Jarod L Roland, Nicholas V Metcalf, Maurizio Corbetta, Gordon L Shulman, and Eric C Leuthardt. Distinct phase-amplitude couplings distinguish cognitive processes in human attention. *NeuroImage*, 175: 111–121, 2018.
- [212] Riitta Hari and Riitta Salmelin. Human cortical oscillations: a neuromagnetic view through the skull. *Trends in neurosciences*, 20(1):44–49, 1997.
- [213] Riitta Salmelin and Riitta Hari. Spatiotemporal characteristics of sensorimotor neuromagnetic rhythms related to thumb movement. *Neuroscience*, 60(2):537–550, 1994.
- [214] Gert Pfurtscheller, Christof Guger, Gernot Müller, Gunther Krausz, and Christa Neuper. Brain oscillations control hand orthosis in a tetraplegic. *Neuroscience letters*, 292(3):211–214, 2000.
- [215] M Alegre, IG Gurtubay, A Labarga, J Iriarte, M Valencia, and J Artieda. Frontal and central oscillatory changes related to different aspects of the motor process: a study in go/no-go paradigms. *Experimental brain research*, 159(1):14–22, 2004.
- [216] Manuel Alegre, Iñaki G Gurtubay, Alberto Labarga, Jorge Iriarte, Armando Malanda, and Julio Artieda. Alpha and beta oscillatory changes during stimulus-induced movement paradigms: effect of stimulus predictability. *Neuroreport*, 14(3):381–385, 2003.
- [217] Craig G Richter, Richard Coppola, and Steven L Bressler. Top-down beta oscillatory signaling conveys behavioral context in early visual cortex. *Scientific reports*, 8(1):1–12, 2018.
- [218] Steven L Bressler and Craig G Richter. Interareal oscillatory synchronization in top-down neocortical processing. *Current opinion in neurobiology*, 31:62–66, 2015.
- [219] William Sedley, Phillip E Gander, Sukhbinder Kumar, Christopher K Kovach, Hiroyuki Oya, Hiroto Kawasaki, III Howard, Matthew A, and Timothy D Griffiths. Neural signatures of perceptual inference. *eLife*, 5:e11476, March 2016. ISSN 2050-084X. doi: 10.7554/eLife.11476. URL <https://doi.org/10.7554/eLife.11476>.
- [220] James J Bonaiuto, Simon Little, Samuel A Neymotin, Stephanie R Jones, Gareth R Barnes, and Sven Bestmann. Laminar dynamics of high amplitude beta bursts in human motor cortex. *Neuroimage*, 242:118479, 2021.

- [221] Maxwell A Sherman, Shane Lee, Robert Law, Saskia Haegens, Catherine A Thorn, Matti S Hämäläinen, Christopher I Moore, and Stephanie R Jones. Neural mechanisms of transient neocortical beta rhythms: Converging evidence from humans, computational modeling, monkeys, and mice. *Proceedings of the National Academy of Sciences*, 113(33):E4885–E4894, 2016.
- [222] David C Knill and Alexandre Pouget. The bayesian brain: the role of uncertainty in neural coding and computation. *TRENDS in Neurosciences*, 27(12):712–719, 2004.
- [223] Markus Bauer, Max-Philipp Stenner, Karl J Friston, and Raymond J Dolan. Attentional modulation of alpha/beta and gamma oscillations reflect functionally distinct processes. *Journal of Neuroscience*, 34(48):16117–16125, 2014.
- [224] Wolfgang Klimesch, Paul Sauseng, and Simon Hanslmayr. EEG alpha oscillations: The inhibition–timing hypothesis. *Brain Research Reviews*, 53(1):63–88, January 2007. ISSN 01650173. doi: 10.1016/j.brainresrev.2006.06.003. URL <http://linkinghub.elsevier.com/retrieve/pii/S016501730600083X>.
- [225] Wolfgang Klimesch. Memory processes, brain oscillations and eeg synchronization. *International journal of psychophysiology*, 24(1-2):61–100, 1996.
- [226] Wolfgang Klimesch, HANNES Schimke, and Gert Pfurtscheller. Alpha frequency, cognitive load and memory performance. *Brain topography*, 5(3):241–251, 1993.
- [227] Arjen M Strijkstra, Domien GM Beersma, Berdine Drayer, Nynke Halbesma, and Serge Daan. Subjective sleepiness correlates negatively with global alpha (8–12 hz) and positively with central frontal theta (4–8 hz) frequencies in the human resting awake electroencephalogram. *Neuroscience letters*, 340(1):17–20, 2003.
- [228] Gert Pfurtscheller. The cortical activation model (cam). *Progress in brain research*, 159:19–27, 2006.
- [229] Karl J Friston, Christian Buechel, Gereon R Fink, Jond Morris, Edmund Rolls, and Raymond J Dolan. Psychophysiological and modulatory interactions in neuroimaging. *Neuroimage*, 6(3):218–229, 1997.
- [230] Michael Woertz, Gert Pfurtscheller, and Wolfgang Klimesch. Alpha power dependent light stimulation: dynamics of event-related (de) synchronization in human electroencephalogram. *Cognitive brain research*, 20(2):256–260, 2004.
- [231] Ole Jensen and Ali Mazaheri. Shaping Functional Architecture by Oscillatory Alpha Activity: Gating by Inhibition. *Frontiers in Human Neuroscience*, 4, 2010. ISSN 1662-5161. doi: 10.3389/fnhum.2010.00186. URL <http://journal.frontiersin.org/article/10.3389/fnhum.2010.00186/abstract>.
- [232] Camille Fakche, Rufin VanRullen, Philippe Marque, and Laura Dugué. α phase-amplitude tradeoffs predict visual perception. *Eneuro*, 9(1), 2022.

- [233] Malte Wöstmann, Viola S Störmer, Jonas Obleser, SK Andersen, N Gaspelin, JJ Geng, SJ Luck, MP Noonan, HA Slagter, J Theeuwes, et al. Ten simple rules to study distractor suppression. *Progress in neurobiology*, page 102269, 2022.
- [234] Michael X Cohen. Where does eeg come from and what does it mean? *Trends in neurosciences*, 40(4):208–218, 2017.
- [235] Y. B. Saalman, M. A. Pinsk, L. Wang, X. Li, and S. Kastner. The Pulvinar Regulates Information Transmission Between Cortical Areas Based on Attention Demands. *Science*, 337(6095):753–756, August 2012. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1223082. URL <http://www.sciencemag.org/cgi/doi/10.1126/science.1223082>.
- [236] Laurie R Silva, Yael Amitai, and Barry W Connors. Intrinsic oscillations of neocortex generated by layer 5 pyramidal neurons. *Science*, 251(4992):432–435, 1991.
- [237] Jorn Niessing, Boris Ebisch, Kerstin E Schmidt, Michael Niessing, Wolf Singer, and Ralf AW Galuske. Hemodynamic signals correlate tightly with synchronized gamma oscillations. *science*, 309(5736):948–951, 2005.
- [238] René Scheeringa, Peter J. Koopmans, Tim van Mourik, Ole Jensen, and David G. Norris. The relationship between oscillatory EEG activity and the laminar-specific BOLD signal. *Proceedings of the National Academy of Sciences*, 113(24):6761–6766, June 2016. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1522577113. URL <http://www.pnas.org/lookup/doi/10.1073/pnas.1522577113>.
- [239] Andre M Bastos, Farran Briggs, Henry J Alitto, George R Mangun, and W Martin Usrey. Simultaneous recordings from the primary visual cortex and lateral geniculate nucleus reveal rhythmic interactions and a cortical source for gamma-band oscillations. *Journal of Neuroscience*, 34(22):7639–7644, 2014.
- [240] Pascal Fries. Rhythms for Cognition: Communication through Coherence. *Neuron*, 88(1): 220–235, October 2015. ISSN 08966273. doi: 10.1016/j.neuron.2015.09.034. URL <https://linkinghub.elsevier.com/retrieve/pii/S0896627315008235>.
- [241] Conrado A Bosman, Jan-Mathijs Schoffelen, Nicolas Brunet, Robert Oostenveld, Andre M Bastos, Thilo Womelsdorf, Birthe Rubehn, Thomas Stieglitz, Peter De Weerd, and Pascal Fries. Attentional stimulus selection through selective synchronization between monkey visual areas. *Neuron*, 75(5):875–888, 2012.
- [242] Georgios Michalareas, Julien Vezoli, Stan van Pelt, Jan-Mathijs Schoffelen, Henry Kennedy, and Pascal Fries. Alpha-Beta and Gamma Rhythms Subserve Feedback and Feedforward Influences among Human Visual Cortical Areas. *Neuron*, 89(2):384–397, January 2016. ISSN 08966273. doi: 10.1016/j.neuron.2015.12.018.
- [243] MARGARET S Livingstone. Oscillatory firing and interneuronal correlations in squirrel monkey striate cortex. *Journal of neurophysiology*, 75(6):2467–2485, 1996.

- [244] Georgia G Gregoriou, Stephen J Gotts, Huihui Zhou, and Robert Desimone. High-frequency, long-range coupling between prefrontal and visual cortex during attention. *science*, 324(5931):1207–1210, 2009.
- [245] Daniel Baldauf and Robert Desimone. Neural mechanisms of object-based attention. *Science*, 344(6182):424–427, 2014.
- [246] Andre M Bastos, Julien Vezoli, and Pascal Fries. Communication through coherence with inter-areal delays. *Current opinion in neurobiology*, 31:173–180, 2015.
- [247] Alexandre Hyafil, Anne-Lise Giraud, Lorenzo Fontolan, and Boris Gutkin. Neural cross-frequency coupling: connecting architectures, mechanisms, and functions. *Trends in neurosciences*, 38(11):725–740, 2015.
- [248] Esther Florin and Sylvain Baillet. The brain’s resting-state activity is shaped by synchronized cross-frequency coupling of neural oscillations. *Neuroimage*, 111:26–35, 2015.
- [249] Hugh Pastoll, Lukas Solanka, Mark CW van Rossum, and Matthew F Nolan. Feedback inhibition enables theta-nested gamma oscillations and grid firing fields. *Neuron*, 77(1):141–154, 2013.
- [250] Douglas McLelland and Rufin VanRullen. Theta-gamma coding meets communication-through-coherence: neuronal oscillatory multiplexing theories reconciled. *PLoS computational biology*, 12(10):e1005162, 2016.
- [251] Tom R Marshall, Jacinta O’Shea, Ole Jensen, and Til O Bergmann. Frontal eye fields control attentional modulation of alpha and gamma oscillations in contralateral occipitoparietal cortex. *Journal of Neuroscience*, 35(4):1638–1647, 2015.
- [252] Freek van Ede, Szabolcs Szebényi, and Eric Maris. Attentional modulations of somatosensory alpha, beta and gamma oscillations dissociate between anticipation and stimulus processing. *Neuroimage*, 97:134–141, 2014.
- [253] Sam M Doesburg, Nicolas Bedo, and Lawrence M Ward. Top-down alpha oscillatory network interactions during visuospatial attention orienting. *Neuroimage*, 132:512–519, 2016.
- [254] Maurizio Corbetta and Gordon L Shulman. Control of goal-directed and stimulus-driven attention in the brain. *Nature reviews neuroscience*, 3(3):201–215, 2002.
- [255] Mathilde Bonnefond and Ole Jensen. Gamma activity coupled to alpha phase as a mechanism for top-down controlled gating. *PloS one*, 10(6):e0128667, 2015.
- [256] Philipp Berens, Georgios A Keliris, Alexander S Ecker, Nikos K Logothetis, and Andreas S Tolias. Feature selectivity of the gamma-band of the local field potential in primate primary visual cortex. *Frontiers in neuroscience*, page 37, 2008.

- [257] Ole Jensen, Jochen Kaiser, and Jean-Philippe Lachaux. Human gamma-frequency oscillations associated with attention and memory. *Trends in neurosciences*, 30(7):317–324, 2007.
- [258] Mathilde Bonnefond, Ole Jensen, and Tommy Clausner. Visual processing by dynamic phase coding. Manuscript submitted for publication, 2022.
- [259] Ole Jensen, Mathilde Bonnefond, and Rufin VanRullen. An oscillatory mechanism for prioritizing salient unattended stimuli. *Trends in cognitive sciences*, 16(4):200–206, 2012.
- [260] Agnes Korcsak-Gorzo, Michael G Müller, Andreas Baumbach, Luziwei Leng, Oliver J Breiwwieser, Sacha J van Albada, Walter Senn, Karlheinz Meier, Robert Legenstein, and Mihai A Petrovici. Cortical oscillations support sampling-based computations in spiking neural networks. *PLoS computational biology*, 18(3):e1009753, 2022.
- [261] Sean M Montgomery, Martha I Betancur, and György Buzsáki. Behavior-dependent coordination of multiple theta dipoles in the hippocampus. *Journal of Neuroscience*, 29(5):1381–1394, 2009.
- [262] Diego Lozano-Soldevilla. On the physiological modulation and potential mechanisms underlying parieto-occipital alpha oscillations. *Frontiers in computational neuroscience*, 12:23, 2018.
- [263] Ben M Harvey, Mariska J Vansteensel, Cyrille H Ferrier, Natalia Petridou, Wietske Zuiderbaan, Erik J Aarnoutse, Martin G Bleichner, H Chris Dijkerman, Martine JE van Zandvoort, Frans SS Leijten, et al. Frequency specific spatial interactions in human electrocorticography: V1 alpha oscillations reflect surround suppression. *Neuroimage*, 65:424–432, 2013.
- [264] Jianrong Jia, Ying Fan, and Huan Luo. Alpha-band phase modulates bottom-up feature processing. *Cerebral Cortex*, 32(6):1260–1268, 2022.
- [265] René Scheeringa, Mathilde Bonnefond, Tim van Mourik, Ole Jensen, David G Norris, and Peter J Koopmans. Relating neural oscillations to laminar fmri connectivity in visual cortex. *Cerebral Cortex*, 2022.
- [266] Sofie S Meyer, James Bonaiuto, Mark Lim, Holly Rossiter, Sheena Waters, David Bradbury, Sven Bestmann, Matthew Brookes, Martina F Callaghan, Nikolaus Weiskopf, et al. Flexible head-casts for high spatial precision meg. *Journal of neuroscience methods*, 276:38–45, 2017.
- [267] James J Bonaiuto, Fardin Afdideh, Maxime Ferez, Konrad Wagstyl, Jérémie Mattout, Mathilde Bonnefond, Gareth R Barnes, and Sven Bestmann. Estimates of cortical column orientation improve meg source inversion. *Neuroimage*, 216:116862, 2020.
- [268] James J Bonaiuto, Sofie S Meyer, Simon Little, Holly Rossiter, Martina F Callaghan, Frederic Dick, Gareth R Barnes, and Sven Bestmann. Lamina-specific cortical dynamics in human visual and sensorimotor cortices. *Elife*, 7:e33977, 2018.

- [269] Marie Simonet, Hadj Boumediene Meziane, Oliver Richard Runswick, Jamie Stephen North, Andrew Mark Williams, Jérôme Barral, and André Roca. The modulation of event-related alpha rhythm during the time course of anticipation. *Scientific Reports*, 9(1):1–11, 2019.
- [270] Mila Halgren, István Ulbert, Hélène Bastuji, Dániel Fabó, Lorand Eröss, Marc Rey, Orrin Devinsky, Werner K Doyle, Rachel Mak-McCully, Eric Halgren, et al. The generation and propagation of the human alpha rhythm. *Proceedings of the National Academy of Sciences*, 116(47):23772–23782, 2019.
- [271] S. Haegens, A. Barczak, G. Musacchia, M. L. Lipton, A. D. Mehta, P. Lakatos, and C. E. Schroeder. Laminar Profile and Physiology of the Rhythm in Primary Visual, Auditory, and Somatosensory Regions of Neocortex. *Journal of Neuroscience*, 35(42):14341–14352, October 2015. ISSN 0270-6474, 1529-2401. doi: 10.1523/JNEUROSCI.0600-15.2015. URL <http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.0600-15.2015>.
- [272] Timo Van Kerkoerle, Matthew W Self, and Pieter R Roelfsema. Layer-specificity in the effects of attention and working memory on activity in primary visual cortex. *Nature communications*, 8(1):1–14, 2017.
- [273] Peter J Koopmans, Markus Barth, and David G Norris. Layer-specific bold activation in human v1. *Human brain mapping*, 31(9):1297–1304, 2010.
- [274] James J. Bonaiuto, Holly E. Rossiter, Sofie S. Meyer, Natalie Adams, Simon Little, Martina F. Callaghan, Fred Dick, Sven Bestmann, and Gareth R. Barnes. Non-invasive laminar inference with MEG: Comparison of methods and source inversion algorithms. *NeuroImage*, 167:372–383, February 2018. ISSN 10538119. doi: 10.1016/j.neuroimage.2017.11.068. URL <https://linkinghub.elsevier.com/retrieve/pii/S1053811917310145>.
- [275] Tzvetan Popov, Ole Jensen, and Jan-Mathijs Schoffelen. Dorsal and ventral cortices are coupled by cross-frequency interactions during working memory. *NeuroImage*, 178:277–286, 2018.
- [276] Ian C Fiebelkorn and Sabine Kastner. Functional specialization in the attention network. *Annual review of psychology*, 71:221, 2020.
- [277] Sabine Kastner, Ian C Fiebelkorn, and Manoj K Eradath. Dynamic pulvino-cortical interactions in the primate attention network. *Current opinion in neurobiology*, 65:10–19, 2020.
- [278] J Ridley Stroop. Studies of interference in serial verbal reactions. *Journal of experimental psychology*, 18(6):643, 1935.
- [279] J Mark G Williams, Andrew Mathews, and Colin MacLeod. The emotional stroop task and psychopathology. *Psychological bulletin*, 120(1):3, 1996.
- [280] Colin M MacLeod. Half a century of research on the stroop effect: an integrative review. *Psychological bulletin*, 109(2):163, 1991.

- [281] Nancy Kanwisher, Josh McDermott, and Marvin M Chun. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of neuroscience*, 17(11):4302–4311, 1997.
- [282] Stanislas Dehaene and Laurent Cohen. The unique role of the visual word form area in reading. *Trends in cognitive sciences*, 15(6):254–262, 2011.
- [283] Melvyn A Goodale and A David Milner. Separate visual pathways for perception and action. *Trends in neurosciences*, 15(1):20–25, 1992.
- [284] Maurizio Capra, Beatrice Bussolino, Alberto Marchisio, Guido Masera, Maurizio Martina, and Muhammad Shafique. Hardware and software optimizations for accelerating deep neural networks: Survey of current trends, challenges, and the road ahead. *IEEE Access*, 8:225134–225180, 2020.
- [285] Masakazu Matsugu, Katsuhiko Mori, Mie Ishii, and Yusuke Mitarai. Convolutional spiking neural network model for robust face detection. In *Proceedings of the 9th International Conference on Neural Information Processing, 2002. ICONIP’02.*, volume 2, pages 660–664. IEEE, 2002.
- [286] Daniel J Saunders, Hava T Siegelmann, Robert Kozma, et al. Stdp learning of image patches with convolutional spiking neural networks. In *2018 international joint conference on neural networks (IJCNN)*, pages 1–7. IEEE, 2018.
- [287] Lei Zhang, Shengyuan Zhou, Tian Zhi, Zidong Du, and Yunji Chen. Tdsnn: From deep neural networks to deep spike neural networks with temporal-coding. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 1319–1326, 2019.
- [288] Gourav Datta and Peter A Beerel. Can deep neural networks be converted to ultra low-latency spiking neural networks? In *2022 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, pages 718–723. IEEE, 2022.
- [289] Charlotte Caucheteux and Jean-Rémi King. Language processing in brains and deep neural networks: computational convergence and its limits. *BioRxiv*, pages 2020–07, 2021.
- [290] Shimon Ullman. Using neuroscience to develop artificial intelligence. *Science*, 363(6428):692–693, 2019.
- [291] Satu Palva and J. Matias Palva. New vistas for α -frequency band oscillations. *Trends in Neurosciences*, 30(4):150–158, April 2007. ISSN 01662236. doi: 10.1016/j.tins.2007.02.001. URL <https://linkinghub.elsevier.com/retrieve/pii/S0166223607000264>.
- [292] Nicholas Gaspelin and Steven J Luck. The role of inhibition in avoiding distraction by salient stimuli. *Trends in cognitive sciences*, 22(1):79–92, 2018.

- [293] Dirk van Moorselaar and Heleen A Slagter. Inhibition in selective attention. *Annals of the New York Academy of Sciences*, 1464(1):204–221, 2020.
- [294] Rodney J. Douglas and Kevan A.C. Martin. NEURONAL CIRCUITS OF THE NEOCORTEX. *Annual Review of Neuroscience*, 27(1):419–451, July 2004. ISSN 0147-006X, 1545-4126. doi: 10.1146/annurev.neuro.27.070203.144152. URL <http://www.annualreviews.org/doi/10.1146/annurev.neuro.27.070203.144152>.
- [295] R. R. Llinas, A. A. Grace, and Y. Yarom. In vitro neurons in mammalian cortical layer 4 exhibit intrinsic oscillatory activity in the 10- to 50-Hz frequency range. *Proceedings of the National Academy of Sciences*, 88(3):897–901, February 1991. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.88.3.897. URL <http://www.pnas.org/cgi/doi/10.1073/pnas.88.3.897>.
- [296] A. Bollimunta, Y. Chen, C. E. Schroeder, and M. Ding. Neuronal Mechanisms of Cortical Alpha Oscillations in Awake-Behaving Macaques. *Journal of Neuroscience*, 28(40):9976–9988, October 2008. ISSN 0270-6474, 1529-2401. doi: 10.1523/JNEUROSCI.2699-08.2008. URL <https://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.2699-08.2008>.
- [297] Julien Fournier, Aman B. Saleem, E. Mika Diamanti, Miles J. Wells, Kenneth D. Harris, and Matteo Carandini. Mouse Visual Cortex Is Modulated by Distance Traveled and by Theta Oscillations. *Current Biology*, 30(19):3811–3817.e6, October 2020. ISSN 09609822. doi: 10.1016/j.cub.2020.07.006. URL <https://linkinghub.elsevier.com/retrieve/pii/S0960982220309969>.
- [298] Hannah Bos, Markus Diesmann, and Moritz Helias. Identifying Anatomical Origins of Coexisting Oscillations in the Cortical Microcircuit. *PLOS Computational Biology*, 12(10):e1005132, October 2016. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1005132. URL <https://dx.plos.org/10.1371/journal.pcbi.1005132>.
- [299] Samuel J.D. Lawrence, Elia Formisano, Lars Muckli, and Floris P. de Lange. Laminar fMRI: Applications for cognitive neuroscience. *NeuroImage*, July 2017. ISSN 10538119. doi: 10.1016/j.neuroimage.2017.07.004. URL <http://linkinghub.elsevier.com/retrieve/pii/S1053811917305724>.
- [300] Matthew W. Self, Timo van Kerkoerle, Rainer Goebel, and Pieter R. Roelfsema. Benchmarking laminar fMRI: Neuronal spiking and synaptic activity during top-down and bottom-up processing in the different layers of cortex. *NeuroImage*, 197:806–817, August 2019. ISSN 10538119. doi: 10.1016/j.neuroimage.2017.06.045. URL <https://linkinghub.elsevier.com/retrieve/pii/S1053811917305177>.
- [301] K.E. Stephan, F.H. Petzschner, L. Kasper, J. Bayer, K.V. Wellstein, G. Stefanics, K.P. Pruessmann, and J. Heinzle. Laminar fMRI and computational theories of brain function. *NeuroImage*, 197:699–706, August 2019. ISSN 10538119. doi: 10.1016/j.neuroimage.2017.11.001. URL <https://linkinghub.elsevier.com/retrieve/pii/S1053811917309084>.

- [302] Jiajia Yang, Laurentius Huber, Yinghua Yu, and Peter A. Bandettini. Linking cortical circuit models to human cognition with laminar fMRI. *Neuroscience & Biobehavioral Reviews*, 128: 467–478, September 2021. ISSN 01497634. doi: 10.1016/j.neubiorev.2021.07.005. URL <https://linkinghub.elsevier.com/retrieve/pii/S0149763421003031>.
- [303] René Scheeringa, Pascal Fries, Karl-Magnus Petersson, Robert Oostenveld, Iris Grothe, David G. Norris, Peter Hagoort, and Marcel C.M. Bastiaansen. Neuronal Dynamics Underlying High- and Low-Frequency EEG Oscillations Contribute Independently to the Human BOLD Signal. *Neuron*, 69(3):572–583, February 2011. ISSN 08966273. doi: 10.1016/j.neuron.2010.11.044. URL <http://linkinghub.elsevier.com/retrieve/pii/S0896627310010810>.
- [304] Marius Schneider, Ana Clara Brogini, Benjamin Dann, Athanasia Tzanou, Cem Uran, Swathi Sheshadri, Hansjörg Scherberger, and Martin Vinck. A mechanism for inter-areal coherence through communication based on connectivity and oscillatory power. *Neuron*, 109(24):4050–4067.e12, December 2021. ISSN 08966273. doi: 10.1016/j.neuron.2021.09.037. URL <https://linkinghub.elsevier.com/retrieve/pii/S0896627321007108>.
- [305] Julien Vezoli, Martin Vinck, Conrado Arturo Bosman, André Moraes Bastos, Christopher Murphy Lewis, Henry Kennedy, and Pascal Fries. Brain rhythms define distinct interaction networks with differential dependence on anatomy. *Neuron*, 109(23):3862–3878.e5, December 2021. ISSN 08966273. doi: 10.1016/j.neuron.2021.09.052. URL <https://linkinghub.elsevier.com/retrieve/pii/S089662732100725X>.
- [306] Yuri B. Saalman and Sabine Kastner. Cognitive and Perceptual Functions of the Visual Thalamus. *Neuron*, 71(2):209–223, July 2011. ISSN 08966273. doi: 10.1016/j.neuron.2011.06.027. URL <https://linkinghub.elsevier.com/retrieve/pii/S0896627311005575>.
- [307] Magor L. Lőrincz, Katalin A. Kékesi, Gábor Juhász, Vincenzo Crunelli, and Stuart W. Hughes. Temporal Framing of Thalamic Relay-Mode Firing by Phasic Inhibition during the Alpha Rhythm. *Neuron*, 63(5):683–696, September 2009. ISSN 08966273. doi: 10.1016/j.neuron.2009.08.012. URL <https://linkinghub.elsevier.com/retrieve/pii/S0896627309006242>.
- [308] Cristin G Welle and Diego Contreras. Sensory-driven and spontaneous gamma oscillations engage distinct cortical circuitry. *Journal of Neurophysiology*, 115(4):1821–1835, 2016.
- [309] René Michel, Laura Dugué, and Niko A Busch. Distinct contributions of alpha and theta rhythms to perceptual and attentional sampling. *European Journal of Neuroscience*, 55(11-12):3025–3039, 2022.
- [310] M Brant-Zawadzki, G D Gillan, and W R Nitz. MP RAGE: a three-dimensional, T1-weighted, gradient-echo sequence—initial experience in the brain. *Radiology*, 182(3):769–775, March 1992. ISSN 0033-8419, 1527-1315. doi: 10.1148/radiology.182.3.1535892. URL <http://pubs.rsna.org/doi/10.1148/radiology.182.3.1535892>.

- [311] B.A. Poser, P.J. Koopmans, T. Witzel, L.L. Wald, and M. Barth. Three dimensional echo-planar imaging at 7 Tesla. *NeuroImage*, 51(1):261–266, May 2010. ISSN 10538119. doi: 10.1016/j.neuroimage.2010.01.108. URL <http://linkinghub.elsevier.com/retrieve/pii/S1053811910001473>.
- [312] Brain Products Inc, GmbH, Munich, Germany. Brain Products, 2018.
- [313] Tommy Clausner, Sarang S. Dalal, and Maité Crespo-García. Photogrammetry-Based Head Digitization for Rapid and Accurate Localization of EEG Electrodes and MEG Fiducial Markers Using a Single Digital SLR Camera. *Frontiers in Neuroscience*, 11, May 2017. ISSN 1662-453X. doi: 10.3389/fnins.2017.00264.
- [314] SR Research. EyeLink 1000 Plus - The Most Flexible Eye Tracker. *SR Research*, 2018.
- [315] Neurobehavioral Systems, Inc. Neurobehavioral Systems, 2018.
- [316] L. Thaler, A.C. Schütz, M.A. Goodale, and K.R. Gegenfurtner. What Is the Best Fixation Target? The Effect of Target Shape on Stability of Fixational Eye Movements. *Vision Research*, 76:31–42, January 2013. ISSN 00426989. doi: 10.1016/j.visres.2012.10.012.
- [317] M. J. Arcaro, S. A. McMains, B. D. Singer, and S. Kastner. Retinotopic Organization of Human Ventral Visual Cortex. *Journal of Neuroscience*, 29(34):10638–10652, August 2009. ISSN 0270-6474, 1529-2401. doi: 10.1523/JNEUROSCI.2807-09.2009.
- [318] David H. Brainard. The Psychophysics Toolbox. *Spatial Vision*, 10:433–436, 1997.
- [319] Brian A Wandell, Suelika Chial, and Benjamin T Backus. Visualization and measurement of the cortical surface. *Journal of cognitive neuroscience*, 12(5):739–752, 2000.
- [320] Ivan Alvarez, De Haas, Benjamin A, Chris A. Clark, Geraint Rees, and D. Samuel Schwarzkopf. Comparing different stimulus configurations for population receptive field mapping in human fMRI. *Frontiers in Human Neuroscience*, 9, 2015. ISSN 1662-5161. doi: 10.3389/fnhum.2015.00096. URL <https://www.frontiersin.org/articles/10.3389/fnhum.2015.00096/full>.
- [321] Ho-Ling Liu and Jia-Hong Gao. An investigation of the impulse functions for the nonlinear BOLD response in functional MRI. *Magnetic Resonance Imaging*, 18(8):931–938, October 2000. ISSN 0730725X. doi: 10.1016/S0730-725X(00)00214-9. URL <http://linkinghub.elsevier.com/retrieve/pii/S0730725X00002149>.
- [322] Kendrick N. Kay, Jonathan Winawer, Aviv Mezer, and Brian A. Wandell. Compressive Spatial Summation in Human Visual Cortex. *Journal of Neurophysiology*, 110(2):481–494, July 2013. ISSN 0022-3077, 1522-1598. doi: 10.1152/jn.00105.2013.

- [323] Brian B. Avants, Nicholas J. Tustison, Gang Song, Philip A. Cook, Arno Klein, and James C. Gee. A reproducible evaluation of ANTs similarity metric performance in brain image registration. *NeuroImage*, 54(3):2033–2044, February 2011. ISSN 10538119. doi: 10.1016/j.neuroimage.2010.09.025. URL <https://linkinghub.elsevier.com/retrieve/pii/S1053811910012061>.
- [324] Robert Oostenveld, Pascal Fries, Eric Maris, and Jan-Mathijs Schoffelen. FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience*, 2011:1–9, 2011. ISSN 1687-5265, 1687-5273. doi: 10.1155/2011/156869.
- [325] Bruce Fischl. FreeSurfer. *NeuroImage*, 62(2):774–781, August 2012. ISSN 10538119. doi: 10.1016/j.neuroimage.2012.01.021.
- [326] Stephen M. Smith, Mark Jenkinson, Mark W. Woolrich, Christian F. Beckmann, Timothy E.J. Behrens, Heidi Johansen-Berg, Peter R. Bannister, Marilena De Luca, Ivana Drobnjak, David E. Flitney, Rami K. Niazy, James Saunders, John Vickers, Yongyue Zhang, Nicola De Stefano, J. Michael Brady, and Paul M. Matthews. Advances in Functional and Structural MR Image Analysis and Implementation as FSL. *NeuroImage*, 23:S208–S219, January 2004. ISSN 10538119. doi: 10.1016/j.neuroimage.2004.07.051.
- [327] Agisoft, St Petersburg, Russia, LLC. Agisoft Metashape. *Educational Edition*, 2014. URL <https://www.agisoft.com/buy/online-store/educational-license/>.
- [328] C Rorden, HO Karnath, and L Bonilha. MRIcron Dicom to Nifti Converter. *Neuroimaging Informatics Tools and Resources Clearinghouse (NITRC)*, 2012.
- [329] Tommy Clausner. MRI Volume Masker 3000 TM, October 2022. URL <https://doi.org/10.5281/zenodo.7211757>.
- [330] Tim Van Mourik, Jan PJM van der Eerden, Pierre-Louis Bazin, and David G Norris. Laminar signal extraction over extended cortical areas by means of a spatial glm. *PloS one*, 14(3): e0212493, 2019.
- [331] K. J. Friston, editor. *Statistical parametric mapping: the analysis of functional brain images*. Elsevier/Academic Press, Amsterdam ; Boston, 1st ed edition, 2007. ISBN 978-0-12-372560-8.
- [332] Washington-University. Workbench, 2018. URL <https://github.com/Washington-University/workbench>.
- [333] B Avants, C Epstein, M Grossman, and J Gee. Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, 12(1):26–41, February 2008. ISSN 13618415. doi: 10.1016/j.media.2007.06.004. URL <https://linkinghub.elsevier.com/retrieve/pii/S1361841507000606>.

- [334] Oscar Esteban, Christopher J. Markiewicz, Ross W. Blair, Craig A. Moodie, A. Ilkay Isik, Asier Erramuzpe, James D. Kent, Mathias Goncalves, Elizabeth DuPre, Madeleine Snyder, Hiroyuki Oya, Satrajit S. Ghosh, Jessey Wright, Joke Durnez, Russell A. Poldrack, and Krzysztof J. Gorgolewski. fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nature Methods*, 16(1):111–116, January 2019. ISSN 1548-7091, 1548-7105. doi: 10.1038/s41592-018-0235-4. URL <http://www.nature.com/articles/s41592-018-0235-4>.
- [335] Liang Wang, Ryan E.B. Mruzek, Michael J. Arcaro, and Sabine Kastner. Probabilistic Maps of Visual Topography in Human Cortex. *Cerebral Cortex*, 25(10):3911–3931, October 2015. ISSN 1047-3211, 1460-2199. doi: 10.1093/cercor/bhu277. URL <https://academic.oup.com/cercor/article-lookup/doi/10.1093/cercor/bhu277>.
- [336] Rebecca Shafee, Randy L. Buckner, and Bruce Fischl. Gray matter myelination of 1555 human brains using partial volume corrected MRI images. *NeuroImage*, 105:473–485, January 2015. ISSN 10538119. doi: 10.1016/j.neuroimage.2014.10.054. URL <http://linkinghub.elsevier.com/retrieve/pii/S1053811914008933>.
- [337] Barry D. Van Veen and Kevin M. Buckley. Beamforming: A Versatile Approach to Spatial Filtering. *IEEE assp magazine*, 5(2):4–24, 1988. URL <http://users.isy.liu.se/en/rt/fredrik/spcourse/beamform.pdf>.
- [338] P. Adjamian, S.F. Worthen, A. Hillebrand, P.L. Furlong, B.A. Chizh, A.R. Hobson, Q. Aziz, and G.R. Barnes. Effective electromagnetic noise cancellation with beamformers and synthetic gradiometry in shielded and partly shielded environments. *Journal of Neuroscience Methods*, 178(1):120–127, March 2009. ISSN 01650270. doi: 10.1016/j.jneumeth.2008.12.006. URL <http://linkinghub.elsevier.com/retrieve/pii/S0165027008006808>.
- [339] Matthew J. Brookes, Karen J. Mullinger, Claire M. Stevenson, Peter G. Morris, and Richard Bowtell. Simultaneous EEG source localisation and artifact rejection during concurrent fMRI by means of spatial filtering. *NeuroImage*, 40(3):1090–1104, April 2008. ISSN 10538119. doi: 10.1016/j.neuroimage.2007.12.030. URL <http://linkinghub.elsevier.com/retrieve/pii/S1053811907011500>.
- [340] Matthew J. Brookes, Jiri Vrba, Karen J. Mullinger, Gerða Björk Geirsdóttir, Winston X. Yan, Claire M. Stevenson, Richard Bowtell, and Peter G. Morris. Source localisation in concurrent EEG/fMRI: Applications at 7T. *NeuroImage*, 45(2):440–452, April 2009. ISSN 10538119. doi: 10.1016/j.neuroimage.2008.10.047. URL <http://linkinghub.elsevier.com/retrieve/pii/S1053811908011609>.
- [341] Johannes Vorwerk, Robert Oostenveld, Maria Carla Piastra, Lilla Magyari, and Carsten H. Wolters. The FieldTrip-SimBio pipeline for EEG forward solutions. *BioMedical Engineering OnLine*, 17(1), December 2018. ISSN 1475-925X. doi: 10.1186/s12938-018-0463-y. URL <https://biomedical-engineering-online.biomedcentral.com/articles/10.1186/s12938-018-0463-y>.

- [342] Maureen Clerc, Alain Dervieux, O Faugeras, R Keriven, Jan Kybic, and Théo Papadopoulo. Comparison of bem and fem methods for the e/meg problem. In *Proceedings of BIOMAG Conference*, 2002.
- [343] J. Vorwerk, M. Clerc, M. Burger, and C. H. Wolters. Comparison of Boundary Element and Finite Element Approaches to the EEG Forward Problem. *Biomedical Engineering / Biomedizinische Technik*, 57(SI-1 Track-O), January 2012. ISSN 1862-278X, 0013-5585. doi: 10.1515/bmt-2012-4152. URL <https://www.degruyter.com/view/j/bmte.2012.57.issue-s1-0/bmt-2012-4152/bmt-2012-4152.xml>.
- [344] S. Baillet, J. C. Mosher, and R. M. Leahy. Electromagnetic brain mapping. *IEEE Signal Processing Magazine*, 18(6):14–30, November 2001. ISSN 1053-5888. doi: 10.1109/79.962275.
- [345] Christoph M Michel, Gregor Thut, Stéphanie Morand, Asaid Khateb, Alan J Pegna, Rolando Grave de Peralta, Sara Gonzalez, Margitta Seeck, and Theodor Landis. Electric source imaging of human brain functions. *Brain Research Reviews*, 36(2):108–118, October 2001. ISSN 0165-0173. doi: 10.1016/S0165-0173(01)00086-8. URL <http://www.sciencedirect.com/science/article/pii/S0165017301000868>.
- [346] D. Slepian. Prolate Spheroidal Wave Functions, Fourier Analysis, and Uncertainty-V: The Discrete Case. *Bell System Technical Journal*, 57(5):1371–1430, May 1978. ISSN 00058580. doi: 10.1002/j.1538-7305.1978.tb02104.x. URL <https://ieeexplore.ieee.org/document/6771595>.
- [347] Rosanne M Van Diepen, John J Foxe, and Ali Mazaheri. The functional role of alpha-band activity in attentional processing: the current zeitgeist and future outlook. *Current opinion in psychology*, 29:229–238, 2019.
- [348] MathWorks. MATLAB, 2021. URL <https://www.mathworks.com/store/>.
- [349] Samuel J. D. Lawrence, Tim van Mourik, Peter Kok, Peter J. Koopmans, David G. Norris, and Floris P. de Lange. Laminar Organization of Working Memory Signals in Human Visual Cortex. *Current Biology*, 28(21):3435–3440.e4, November 2018. ISSN 0960-9822. doi: 10.1016/j.cub.2018.08.043. URL [https://www.cell.com/current-biology/abstract/S0960-9822\(18\)31127-8](https://www.cell.com/current-biology/abstract/S0960-9822(18)31127-8). Publisher: Elsevier.
- [350] Samuel JD Lawrence, David G Norris, and Floris P de Lange. Dissociable laminar profiles of concurrent bottom-up and top-down modulation in the human visual cortex. *eLife*, 8:e44422, May 2019. ISSN 2050-084X. doi: 10.7554/eLife.44422. URL <https://doi.org/10.7554/eLife.44422>. Publisher: eLife Sciences Publications, Ltd.
- [351] Irati Markuerkiaga, José P Marques, Lauren J Bains, and David G Norris. An in-vivo study of bold laminar responses as a function of echo time and static magnetic field strength. *Scientific reports*, 11(1):1–13, 2021.

- [352] Eric Maris and Robert Oostenveld. Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164(1):177–190, August 2007. ISSN 0165-0270. doi: 10.1016/j.jneumeth.2007.03.024. URL <https://www.sciencedirect.com/science/article/pii/S0165027007001707>.
- [353] Tommy Clausner and Stefano Gentili. Auto-regressive Rank Order Similarity (aros) test, June 2022. URL <https://www.biorxiv.org/content/10.1101/2022.06.15.496113v1>. Pages: 2022.06.15.496113 Section: New Results.
- [354] Roger BH Tootell, Nouchine Hadjikhani, E Kevin Hall, Sean Marrett, Wim Vanduffel, J Thomas Vaughan, and Anders M Dale. The retinotopy of visual spatial attention. *Neuron*, 21(6):1409–1422, 1998.
- [355] Sofia Crespi, Laura Biagi, Giovanni d’Avossa, David C Burr, Michela Tosetti, and Maria Concetta Morrone. Spatiotopic coding of bold signal in human visual cortex depends on spatial attention. *PLoS one*, 6(7):e21661, 2011.
- [356] Dardo Tomasi, Thomas Ernst, Elisabeth C Caparelli, and Linda Chang. Common deactivation patterns during working memory and visual attention tasks: An intra-subject fmri study at 4 tesla. *Human brain mapping*, 27(8):694–705, 2006.
- [357] Benjamin Y Hayden and Jack L Gallant. Time course of attention reveals different mechanisms for spatial and feature-based attention in area v4. *Neuron*, 47(5):637–643, 2005.
- [358] Pei Sun, Justin L Gardner, Mauro Costagli, Kenichi Ueno, R Allen Waggoner, Keiji Tanaka, and Kang Cheng. Demonstration of tuning to stimulus orientation in the human visual cortex: a high-resolution fmri study with a novel continuous and periodic stimulation paradigm. *Cerebral Cortex*, 23(7):1618–1629, 2013.
- [359] Stephen A Engel, Gary H Glover, and Brian A Wandell. Retinotopic organization in human visual cortex and the spatial precision of functional mri. *Cerebral cortex (New York, NY: 1991)*, 7(2):181–192, 1997.
- [360] Lars Michels, Kerstin Bucher, Rafael Luchinger, Peter Klaver, Ernst Martin, Daniel Jeanmonod, and Daniel Brandeis. Simultaneous eeg-fmri during a working memory task: modulations in low and high frequency bands. *PLoS one*, 5(4):e10298, 2010.
- [361] Makoto Uji, Ross Wilson, Susan T Francis, Karen J Mullinger, and Stephen D Mayhew. Exploring the advantages of multiband fmri with simultaneous eeg to investigate coupling between gamma frequency neural activity and the bold response in humans. *Human Brain Mapping*, 39(4):1673–1687, 2018.
- [362] Ángel Correa, Juan Lupiáñez, Eduardo Madrid, and Pío Tudela. Temporal attention enhances early visual processing: A review and new evidence from event-related potentials. *Brain research*, 1076(1):116–128, 2006.

- [363] Behzad Zareian, Kouros Maboudi, Mohammad Reza Daliri, Hamid Abrishami Moghaddam, Stefan Treue, and Moein Esghaei. Attention strengthens across-trial pre-stimulus phase coherence in visual cortex, enhancing stimulus processing. *Scientific reports*, 10(1):1–12, 2020.
- [364] Nikos K Logothetis, Jon Pauls, Mark Augath, Torsten Trinath, and Axel Oeltermann. Neurophysiological investigation of the basis of the fmri signal. *nature*, 412(6843):150–157, 2001.
- [365] György Buzsáki, Costas A Anastassiou, and Christof Koch. The origin of extracellular fields and currents—eeg, ecog, lfp and spikes. *Nature reviews neuroscience*, 13(6):407–420, 2012.
- [366] Simon Musall, Veronika Von Pfössl, Alexander Rauch, Nikos K Logothetis, and Kevin Whittingstall. Effects of neural synchrony on surface eeg. *Cerebral Cortex*, 24(4):1045–1053, 2014.
- [367] Marco Bimbi, Fabrizia Festante, Gino Coudé, Ross E Vanderwert, Nathan A Fox, and Pier Francesco Ferrari. Simultaneous scalp recorded eeg and local field potentials from monkey ventral premotor cortex during action observation and execution reveals the contribution of mirror and motor neurons to the mu-rhythm. *Neuroimage*, 175:22–31, 2018.
- [368] Chuanliang Han, Tian Wang, Yi Yang, Yujie Wu, Yang Li, Weifeng Dai, Yange Zhang, Bin Wang, Guanzhong Yang, Ziqi Cao, et al. Multiple gamma rhythms carry distinct spatial frequency information in primary visual cortex. *PLoS Biology*, 19(12):e3001466, 2021.
- [369] Xiaoxuan Jia, Seiji Tanabe, and Adam Kohn. Gamma and the coordination of spiking activity in early visual cortex. *Neuron*, 77(4):762–774, 2013.
- [370] René Scheeringa, Karl Magnus Petersson, Robert Oostenveld, David G Norris, Peter Hagoort, and Marcel CM Bastiaansen. Trial-by-trial coupling between eeg and bold identifies networks related to alpha and theta eeg power increases during working memory maintenance. *Neuroimage*, 44(3):1224–1238, 2009.
- [371] Lars Michels, Rafael Lüchinger, Thomas Koenig, Ernst Martin, and Daniel Brandeis. Developmental changes of bold signal correlations with global human eeg power and synchronization during working memory. *PLoS One*, 7(7):e39447, 2012.
- [372] Helmut Laufs, Andreas Kleinschmidt, Astrid Beyerle, Evelyn Eger, Afraim Salek-Haddadi, Christine Preibisch, and Karsten Krakow. Eeg-correlated fmri of human alpha activity. *Neuroimage*, 19(4):1463–1476, 2003.
- [373] Manbir Singh, Sungheon Kim, and Tae-Seong Kim. Correlation between bold-fmri and eeg signal changes in response to visual stimulus frequency in humans. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 49(1):108–114, 2003.
- [374] Petra Ritter, Matthias Moosmann, and Arno Villringer. Rolandic alpha and beta eeg rhythms’ strengths are inversely related to fmri-bold signal in primary somatosensory and motor cortex. *Human brain mapping*, 30(4):1168–1187, 2009.

- [375] Simon Hanslmayr, Gregor Volberg, Maria Wimber, Markus Raabe, Mark W Greenlee, and Karl-Heinz T Bäuml. The relationship between brain oscillations and bold signal during memory formation: a combined eeg–fmri study. *Journal of Neuroscience*, 31(44):15674–15680, 2011.
- [376] Stephen D Mayhew, Dirk Ostwald, Camillo Porcaro, and Andrew P Bagshaw. Spontaneous eeg alpha oscillation interacts with positive and negative bold responses in the visual–auditory cortices and default-mode network. *Neuroimage*, 76:362–372, 2013.
- [377] Kyle E Mathewson, Alejandro Lleras, Diane M Beck, Monica Fabiani, Tony Ro, and Gabriele Gratton. Pulsed out of awareness: Eeg alpha oscillations represent a pulsed-inhibition of ongoing cortical processing. *Frontiers in psychology*, 2:99, 2011.
- [378] Michael S Clayton, Nick Yeung, and Roi Cohen Kadosh. The roles of cortical oscillations in sustained attention. *Trends in cognitive sciences*, 19(4):188–195, 2015.
- [379] Mattia F Pagnotta, David Pascucci, and Gijs Plomp. Nested oscillations and brain connectivity during sequential stages of feature-based attention. *NeuroImage*, 223:117354, 2020.
- [380] Daeun Gwon and Minkyu Ahn. Alpha and high gamma phase amplitude coupling during motor imagery and weighted cross-frequency coupling to extract discriminative cross-frequency patterns. *NeuroImage*, 240:118403, 2021.
- [381] Michael D Ernst. Permutation methods: a basis for exact inference. *Statistical Science*, pages 676–685, 2004.
- [382] Marti Anderson and Cajo Ter Braak. Permutation tests for multi-factorial analysis of variance. *Journal of statistical computation and simulation*, 73(2):85–113, 2003.
- [383] Douglas M Potter. A permutation test for inference in logistic regression with small-and moderate-sized data sets. *Statistics in medicine*, 24(5):693–708, 2005.
- [384] Kenneth J Berry, Janis E Johnston, and Paul W Mielke Jr. Permutation methods. *Wiley Interdisciplinary Reviews: Computational Statistics*, 3(6):527–542, 2011.
- [385] Anders Odén and Hans Wedel. Arguments for fisher’s permutation test. *The Annals of Statistics*, pages 518–520, 1975.
- [386] Rick Durrett. *Probability: theory and examples*, volume 49. Cambridge university press, 2019. ISBN 9780198520115.
- [387] Herbert A Sturges. The choice of a class interval. *Journal of the american statistical association*, 21(153):65–66, 1926.
- [388] David Freedman and Persi Diaconis. On the histogram as a density estimator: L 2 theory. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 57(4):453–476, 1981.

- [389] numpy. `numpy.histogram_bin_edges`. https://numpy.org/doc/stable/reference/generated/numpy.histogram_bin_edges.html, 2022. [Online; accessed 01-June-2022].
- [390] Irati Markuerkiaga, José P Marques, Lauren J Bains, and David G Norris. An in-vivo study of bold laminar responses as a function of echo time and static magnetic field strength. *Scientific reports*, 11(1):1–13, 2021.
- [391] Stephen C Strother. Evaluating fmri preprocessing pipelines. *IEEE Engineering in Medicine and Biology Magazine*, 25(2):27–41, 2006.
- [392] Tom Johnstone, Kathleen S Ores Walsh, Larry L Greischar, Andrew L Alexander, Andrew S Fox, Richard J Davidson, and Terrence R Oakes. Motion correction and the use of motion covariates in multiple-subject fmri analysis. *Human brain mapping*, 27(10):779–788, 2006.
- [393] René Scheeringa and Pascal Fries. Cortical layers, rhythms and bold signals. *NeuroImage*, 197:689–698, 2019.
- [394] dipy. `dipy.segment.mask.median_otsu`. https://dipy.org/documentation/1.0.0./examples_built/brain_extraction_dwi/, 2022. [Online; accessed 12-October-2022].
- [395] Krzysztof J Gorgolewski, Tibor Auer, Vince D Calhoun, R Cameron Craddock, Samir Das, Eugene P Duff, Guillaume Flandin, Satrajit S Ghosh, Tristan Glatard, Yaroslav O Halchenko, et al. The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific data*, 3(1):1–9, 2016.
- [396] Cyril R Perneta, Stefan Appelhoffb, Guillaume Flandinc, Christophe Phillipsd, Arnaud Delorme, and Robert Oostenveldg. Bids-eeg: an extension to the brain imaging data structure (bids) specification for electroencephalography. 2019.
- [397] Guiomar Niso, Krzysztof J Gorgolewski, Elizabeth Bock, Teon L Brooks, Guillaume Flandin, Alexandre Gramfort, Richard N Henson, Mainak Jas, Vladimir Litvak, Jeremy T Moreau, et al. Meg-bids, the brain imaging data structure extended to magnetoencephalography. *Scientific data*, 5(1):1–5, 2018.
- [398] Saskia Haegens, Barbara F Händel, and Ole Jensen. Top-down controlled alpha band activity in somatosensory areas determines behavioral performance in a discrimination task. *Journal of Neuroscience*, 31(14):5197–5204, 2011.
- [399] Yuka O Okazaki, Peter De Weerd, Saskia Haegens, and Ole Jensen. Hemispheric lateralization of posterior alpha reduces distracter interference during face matching. *Brain research*, 1590: 56–64, 2014.
- [400] Chengrou Lu, Huiling Li, Ruilin Fu, Jing Qu, Qingxin Yue, and Leilei Mei. Neural representation in visual word form area during word reading. *Neuroscience*, 452:49–62, 2021.

- [401] Lang Chen, Demian Wassermann, Daniel A Abrams, John Kochalka, Guillermo Gallardo-Diez, and Vinod Menon. The visual word form area (vwfa) is part of both language and attention circuitry. *Nature communications*, 10(1):1–12, 2019.
- [402] Maurizio Corbetta, Gaurav Patel, and Gordon L Shulman. The reorienting system of the human brain: from environment to theory of mind. *Neuron*, 58(3):306–324, 2008.
- [403] Richard C Oldfield. The assessment and analysis of handedness: the edinburgh inventory. *Neuropsychologia*, 9(1):97–113, 1971.
- [404] Nancy Kanwisher and Galit Yovel. The fusiform face area: a cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 361(1476):2109–2128, 2006.
- [405] Qiong Cao, Li Shen, Weidi Xie, Omkar M Parkhi, and Andrew Zisserman. Vggface2: A dataset for recognising faces across pose and age. In *2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)*, pages 67–74. IEEE, 2018.
- [406] Bobby Stojanoski and Rhodri Cusack. Time to wave good-bye to phase scrambling: Creating controlled scrambled images using diffeomorphic transformations. *Journal of vision*, 14(12):6–6, 2014.
- [407] Maxime Ferez. *Functional Role and Top Down Control of Alpha Oscillations*. PhD thesis, University Claude Bernard Lyon 1, 2022.
- [408] J Richard Simon and James D Wolf. Choice reaction time as a function of angular stimulus-response correspondence and age. *Ergonomics*, 6(1):99–105, 1963.
- [409] Marcel CM Bastiaansen and Thomas R Knösche. Tangential derivative mapping of axial meg applied to event-related desynchronization research. *Clinical Neurophysiology*, 111(7):1300–1305, 2000.
- [410] Gregor Thut, Annika Nietzel, Stephan A Brandt, and Alvaro Pascual-Leone. α -band electroencephalographic activity over occipital cortex indexes visuospatial attention bias and predicts visual target detection. *Journal of Neuroscience*, 26(37):9494–9502, 2006.
- [411] W Pieter Medendorp, Geerten FI Kramer, Ole Jensen, Robert Oostenveld, Jan-Mathijs Schoffelen, and Pascal Fries. Oscillatory activity in human parietal and occipital cortex shows hemispheric lateralization and memory effects in a delayed double-step saccade task. *Cerebral cortex*, 17(10):2364–2374, 2007.
- [412] Michael S Worden, John J Foxe, Norman Wang, and Gregory V Simpson. Anticipatory biasing of visuospatial attention indexed by retinotopically specific α -band electroencephalography increases over occipital cortex. *Journal of Neuroscience*, 20(6):RC63–RC63, 2000.

- [413] Wolfgang Klimesch. Alpha-band oscillations, attention, and controlled access to stored information. *Trends in cognitive sciences*, 16(12):606–617, 2012.
- [414] Suresh D Muthukumaraswamy and Krish D Singh. Visual gamma oscillations: the effects of stimulus type, visual field coverage and stimulus motion on meg and eeg recordings. *Neuroimage*, 69:223–230, 2013.
- [415] Catherine Tallon-Baudry, Olivier Bertrand, Marie-Anne Hénaff, Jean Isnard, and Catherine Fischer. Attention modulates gamma-band oscillations differently in the human lateral occipital cortex and fusiform gyrus. *Cerebral cortex*, 15(5):654–662, 2005.
- [416] Wolfgang Klimesch, Michael Doppelmayr, Harald Russegger, Thomas Pachinger, and Jens Schwaiger. Induced alpha band power changes in the human eeg and attention. *Neuroscience letters*, 244(2):73–76, 1998.
- [417] NS Ermachenko, AA Ermachenko, and AV Latanov. Desynchronization α frequency event-related in visual selective attention requiring tasks. *Human Physiology*, 37(6):658–666, 2011.
- [418] Pierre Perruchet and Sebastien Pacton. Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in cognitive sciences*, 10(5):233–238, 2006.
- [419] Anthony G Greenwald, Debbie E McGhee, and Jordan LK Schwartz. Measuring individual differences in implicit cognition: the implicit association test. *Journal of personality and social psychology*, 74(6):1464, 1998.
- [420] Stefanie Liebe, Gregor M Hoerzer, Nikos K Logothetis, and Gregor Rainer. Theta coupling between v4 and prefrontal cortex predicts visual short-term memory performance. *Nature neuroscience*, 15(3):456–462, 2012.
- [421] James F Cavanagh and Michael J Frank. Frontal theta as a mechanism for cognitive control. *Trends in cognitive sciences*, 18(8):414–421, 2014.
- [422] Stan van Pelt, Lieke Heil, Johan Kwisthout, Sasha Ondobaka, Iris van Rooij, and Harold Bekkering. Beta-and gamma-band activity reflect predictive coding in the processing of causal events. *Social cognitive and affective neuroscience*, 11(6):973–980, 2016.
- [423] Joachim Gross, Jan Kujala, Matti Hämäläinen, Lars Timmermann, Alfons Schnitzler, and Riitta Salmelin. Dynamic imaging of coherent sources: studying neural interactions in the human brain. *Proceedings of the National Academy of Sciences*, 98(2):694–699, 2001.
- [424] Guido Nolte. The magnetic lead field theorem in the quasi-static approximation and its use for magnetoencephalography forward calculation in realistic volume conductors. *Physics in Medicine & Biology*, 48(22):3637, 2003.

- [425] Alan C Evans, D Louis Collins, SR Mills, Edward D Brown, Ryan L Kelly, and Terry M Peters. 3d statistical neuroanatomical models from 305 mri volumes. In *1993 IEEE conference record nuclear science symposium and medical imaging conference*, pages 1813–1817. IEEE, 1993.
- [426] Samuel W Greenhouse and Seymour Geisser. On methods in the analysis of profile data. *Psychometrika*, 24(2):95–112, 1959.
- [427] Kenji Matsumoto and Keiji Tanaka. Conflict and cognitive control. *Science*, 303(5660):969–970, 2004.
- [428] Maxime Ferez, Tommy Clausner, Julia Lukacs, Mélinda Gbadoe, Sébastien Daligault, Denis Schwartz, and Mathilde Bonnefond. Functional inhibition of high-order visual regions by alpha oscillations. in prep.
- [429] Gordon D Logan, N Jane Zbrodoff, and James Williamson. Strategies in the color-word stroop task. *Bulletin of the Psychonomic Society*, 22(2):135–138, 1984.
- [430] Gordon M Redding and Deborah A Gerjets. Stroop effect: Interference and facilitation with verbal and manual responses. *Perceptual and Motor Skills*, 45(1):11–17, 1977.
- [431] Alexander Zhigalov and Ole Jensen. Alpha oscillations do not implement gain control in early visual cortex but rather gating in parieto-occipital regions. *Human Brain Mapping*, 41(18):5176–5186, 2020.
- [432] Dirk van Moorselaar and Heleen A Slagter. Learning what is irrelevant or relevant: Expectations facilitate distractor inhibition and target facilitation through distinct neural mechanisms. *Journal of Neuroscience*, 39(35):6953–6967, 2019.
- [433] Elisabet Alzueta, María Melcón, Ole Jensen, and Almudena Capilla. The ‘narcissus effect’: Top-down alpha-beta band modulation of face-related brain areas during self-face processing. *NeuroImage*, 213:116754, 2020.
- [434] Tzvetan Popov, Gregory A Miller, Brigitte Rockstroh, and Nathan Weisz. Modulation of α power and functional connectivity during facial affect recognition. *Journal of Neuroscience*, 33(14):6018–6026, 2013.
- [435] Lin Wang, Peter Hagoort, and Ole Jensen. Language prediction is reflected by coupling between frontal gamma and posterior alpha oscillations. *Journal of cognitive neuroscience*, 30(3):432–447, 2018.
- [436] Jörg Bahlmann, Esther Aarts, and Mark D’Esposito. Influence of motivation on control hierarchy in the human frontal cortex. *Journal of Neuroscience*, 35(7):3207–3217, 2015.
- [437] Raghavendra Kotikalapudi and contributors. keras-vis. <https://github.com/raghakot/keras-vis>, 2017.

- [438] Ying Shang, Yongli Li, Feng You, and RuiLian Zhao. Conversion-based approach to obtain an snn construction. *International Journal of Software Engineering and Knowledge Engineering*, 30(11n12):1801–1818, 2020.
- [439] Gustavo Deco, Viktor K Jirsa, Peter A Robinson, Michael Breakspear, and Karl Friston. The dynamic brain: from spiking neurons to neural masses and cortical fields. *PLoS computational biology*, 4(8):e1000092, 2008.
- [440] Kukan Selvaratnam, Yasuaki Kuroe, and Takehiro Mori. Synthesis of spiking neural oscillators. In *IEEE SMC'99 Conference Proceedings. 1999 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No. 99CH37028)*, volume 1, pages 490–495. IEEE, 1999.
- [441] Edward Wallace, Marc Benayoun, Wim Van Drongelen, and Jack D Cowan. Emergent oscillations in networks of stochastic spiking neurons. *Plos one*, 6(5):e14804, 2011.
- [442] Nikola K Kasabov. Neucube: A spiking neural network architecture for mapping, learning and understanding of spatio-temporal brain data. *Neural Networks*, 52:62–76, 2014.
- [443] Xiaochun Gu, Fang Han, Zhijie Wang, and Xia Peng. Dependency of gamma oscillations in e/i neuronal network on illumination contrast of external stimulus. *Theoretical and Applied Mechanics Letters*, 9(1):14–20, 2019.
- [444] Roosa Honkanen, Santeri Rouhinen, Sheng H Wang, J Matias Palva, and Satu Palva. Gamma oscillations underlie the maintenance of feature-specific information and the contents of visual working memory. *Cerebral cortex*, 25(10):3788–3801, 2015.
- [445] Kyle R Cave and Narcisse P Bichot. Visuospatial attention: Beyond a spotlight model. *Psychonomic bulletin & review*, 6(2):204–223, 1999.
- [446] Ian C Fiebelkorn and Sabine Kastner. A rhythmic theory of attention. *Trends in cognitive sciences*, 23(2):87–101, 2019.
- [447] Masataka Watanabe, Kang Cheng, Yusuke Murayama, Kenichi Ueno, Takeshi Asamizuya, Keiji Tanaka, and Nikos Logothetis. Attention but not awareness modulates the bold signal in the human v1 during binocular suppression. *Science*, 334(6057):829–831, 2011.
- [448] Jaap Munneke, Dirk J Heslenfeld, and Jan Theeuwes. Directing attention to a location in space results in retinotopic activation in primary visual cortex. *Brain research*, 1222:184–191, 2008.
- [449] Andreas Wutz, David Melcher, and Jason Samaha. Frequency modulation of neural oscillations according to visual task demands. *Proceedings of the National Academy of Sciences*, 115(6):1346–1351, 2018.
- [450] Tobias Feldmann-Wüstefeld and Edward Awh. Alpha-band activity tracks the zoom lens of attention. *Journal of cognitive neuroscience*, 32(2):272–282, 2020.

- [451] Catie Chang, Zhongming Liu, Michael C Chen, Xiao Liu, and Jeff H Duyn. Eeg correlates of time-varying bold functional connectivity. *Neuroimage*, 72:227–236, 2013.
- [452] Wolfgang Klimesch. The frequency architecture of brain and brain body oscillations: an analysis. *European Journal of Neuroscience*, 48(7):2431–2453, 2018.
- [453] Ana Navarrete, Carel P Van Schaik, and Karin Isler. Energetics and the evolution of human brain size. *Nature*, 480(7375):91–93, 2011.
- [454] Ravid Shwartz-Ziv and Naftali Tishby. Opening the black box of deep neural networks via information. *arXiv preprint arXiv:1703.00810*, 2017.
- [455] K Seeliger, L Ambrogioni, U Güçlü, and M van Gerven. Neural information flow: Learning neural information processing systems from brain activity. *BioRxiv*, 2019.
- [456] Irwin Sobel, Gary Feldman, et al. A 3x3 isotropic gradient operator for image processing. *a talk at the Stanford Artificial Project in*, pages 271–272, 1968.
- [457] Michael Plöchl, Ian Fiebelkorn, Sabine Kastner, and Jonas Obleser. Attentional sampling of visual and auditory objects is captured by theta-modulated neural activity. *European Journal of Neuroscience*, 55(11-12):3067–3082, 2022.
- [458] Benjamin Peters, Benjamin Rahm, Jochen Kaiser, and Christoph Bledowski. Attention samples objects held in working memory at a theta rhythm. *Preprint at [https://www. biorxiv. org/content/early/2018/07/24/369652](https://www.biorxiv.org/content/early/2018/07/24/369652)*, 2018.
- [459] John O’Keefe. Hippocampus, theta, and spatial memory. *Current opinion in neurobiology*, 3(6):917–924, 1993.
- [460] Katerina Kalemaki, Xanthippi Konstantoudaki, Simona Tivodar, Kyriaki Sidiropoulou, and Domna Karageos. Mice with decreased number of interneurons exhibit aberrant spontaneous and oscillatory activity in the cortex. *Frontiers in neural circuits*, 12:96, 2018.
- [461] Jesse C Werth, Matthew Einhorn, and Thomas A Cleland. Dynamics of spike time encoding in the olfactory bulb. *bioRxiv*, 2022.
- [462] Antonio Fernández-Ruiz, Azahara Oliva, Marisol Soula, Florbela Rocha-Almeida, Gergo A Nagy, Gonzalo Martin-Vazquez, and György Buzsáki. Gamma rhythm communication between entorhinal cortex and dentate gyrus neuronal assemblies. *Science*, 372(6537):eabf3119, 2021.

"In git we trust."