

Decoding implicit information
from the electroencephalogram
with methods from
brain-computer interfacing

vorgelegt von
Dipl. Biol.
Markus Wenzel
aus Heidelberg

von der Fakultät IV – Elektrotechnik und Informatik
der Technischen Universität Berlin

zur Erlangung des akademischen Grades

Doktor der Naturwissenschaften

– Dr. rer. nat. –

genehmigte Dissertation

Promotionsausschuss:

Vorsitzender: Prof. Dr. Henning Sprekeler

Gutachter: Prof. Dr. Benjamin Blankertz

Gutachter: Prof. Dr. Klaus-Robert Müller

Gutachter: Prof. Dr. Peter Desain

Tag der wissenschaftlichen Aussprache: 19. Juni 2017

Berlin, 2017

Wer kann sie erraten?

[Traditional, 1842]

Acknowledgements

I feel exceptionally fortunate for having had the opportunity to learn from Prof. Dr. Benjamin Blankertz and Prof. Dr. Klaus-Robert Müller, and want to express my sincere thanks for great inspiration, guidance and support! I believe that the cheerful and productive atmosphere that Benjamin and Klaus have created is unique and nothing else than a real gift.

The insightful and sharp-witted discussions with Prof. Dr. Gabriel Curio were extremely valuable, as well as the wonderful cooperation with Inês Almeida, Jan-Eike Golenia, Mihail Bogojeski, Frank Meinecke, Fabien Cardinaux and Rafael Schultze-Kraft! Dominik Kühne and Imke Weitkamp were always there with great support! Thank you so much for the awesome time, Alexander, Bastian, Daniel, Duncan, Felix, Han-Jeong, Hendrik, Irene, Irina, Johannes, Laura, Manon, Marija, Martijn, Matthias, Siamac, Stefan, Stephanie and Sven!

M. W.

Abstract

Background. Research on brain-computer interfacing (BCI) has demonstrated that specific brain activity patterns can be detected in the electroencephalogram (EEG) with multivariate methods from machine learning and signal processing in real-time. *Objective.* This direct access to the neural processes potentially provides an opportunity to learn about the users of technical applications in a novel way. In this dissertation, it was explored how implicit, user-related information can be decoded from the EEG. *Approach and results.* First, it was demonstrated that BCI methods can uncover an imperceptible usability flaw of a technical device. A neural workload, which is potentially uncomfortable on the long-term, was imposed by the device on the brain of the user, but could not be noticed by test persons due to the limits of human perception. The findings suggest a remedy, which may improve the ease of use of the assessed device. In a second line of research, it was investigated how the subjective relevance of viewed items can be estimated based on EEG and eye tracking signals. This information renders it possible to map the relevance of the visual surrounding, and to infer the current interest of the individual user in real-time. *Significance.* Signals originating in the brain contain valuable information about the users of technical devices and applications. It was demonstrated that the direct observation of the neural processes offers particular benefits in comparison to standard methods for obtaining user-related information, such as questionnaires. Multivariate methods proved to be essential for extracting information about the complex neural activity from the recorded signals. The methods recognised patterns that were distributed over the numerous dimensions of the EEG data and obscured by irrelevant activity and noise.

Keywords: Brain-computer interfacing, electroencephalography, eye tracking, neuroergonomics, neurotechnology, pattern recognition, relevance detection

Zusammenfassung

Hintergrund. Forschung auf dem Gebiet der Gehirn-Computer-Schnittstelle hat gezeigt, dass spezifische Gehirnaktivitätsmuster mit multivariaten Methoden des maschinellen Lernens und der Signalverarbeitung im Elektroenzephalogramm (EEG) in Echtzeit erkannt werden können. *Ziel.* Mit diesem direkten Zugang zu den neuronalen Prozessen kann man möglicherweise etwas über die Nutzer technischer Anwendungen auf neuartige Weise erfahren. In dieser Dissertation wurde untersucht, wie implizite, nutzerbezogene Information aus dem EEG entschlüsselt werden kann. *Ansatz und Ergebnisse.* Zunächst wurde mit der Gehirn-Computer-Schnittstelle ein Mangel der Nutzbarkeit eines technischen Gerätes aufgedeckt. Das Gerät strengte das Gehirn des Nutzers auf nicht wahrnehmbare Weise an, was auf lange Sicht unangenehm sein könnte. Testpersonen konnten den untersuchten Mangel aufgrund der Grenzen der menschlichen Wahrnehmung nicht bemerken. Die Ergebnisse geben einen Hinweis darauf, wie diese unnötige Beanspruchung möglicherweise vermieden werden könnte. Desweiteren wurde untersucht, wie die subjektive Relevanz von betrachteten Objekten anhand des EEG und der Augenbewegungen eingeschätzt werden kann. Diese Information erlaubt es, die Relevanz der visuellen Umgebung zu kartieren und in Echtzeit Rückschlüsse auf das Interesse des individuellen Nutzers zu ziehen. *Bedeutung.* Messungen der Gehirnaktivität enthalten wertvolle Information über die Nutzer technischer Geräte und Anwendungen. Es wurde gezeigt, dass die direkte Beobachtung neuronaler Prozesse besondere Vorteile im Vergleich zu Standardverfahren zur Gewinnung nutzerbezogener Information bietet, wie etwa Fragebögen. Multivariate Methoden erwiesen sich als wesentlich, um Information über die komplexe neuronale Aktivität aus den Signalen zu extrahieren. Die Methoden erkannten Muster, die über die zahlreichen Dimensionen der EEG Daten verteilt und durch irrelevante Aktivität und Rauschen überlagert waren.

Stichwörter: Gehirn-Computer-Schnittstelle, Elektroenzephalographie, Augenbewegungen, Neuroergonomie, Neurotechnologie, Mustererkennung, Relevanzerkennung

Contents

Acknowledgements	i
Abstract (English/Deutsch)	iii
List of figures	xi
List of tables	xiii
1 Introduction	1
1.1 Overview	1
1.2 Outline of the dissertation	2
1.3 List of included publications	4
1.4 Additional publications	4
1.5 Conference contributions	5
2 Fundamentals	7
2.1 Measuring neural processes in the human brain	7
2.2 Brain-computer interfacing and mental state decoding	9
2.2.1 Brain-computer interfacing for communication and control	10
2.2.2 Mental state decoding	12
2.3 Multivariate methods for detecting information in the EEG	13
2.3.1 Linear generative model of the EEG	14
2.3.2 Demixing the EEG	15
2.3.3 Recognising patterns in the EEG	16
2.4 Lessons learned	24
3 EEG-based usability assessment of stereoscopic displays	25
3.1 Introduction	25
3.2 Material and methods	27
3.2.1 Experimental design	27
3.2.2 Selection of the shutter frequencies	27
3.2.3 Determination of the individual flicker fusion threshold	28
3.2.4 Data acquisition	29
3.2.5 Experimental setup	29
3.2.6 Data analysis	30

Contents

3.3	Results	32
3.3.1	Flicker fusion thresholds and shutter frequencies	32
3.3.2	ICA to extract neural correlates of the flicker	32
3.3.3	Quantification of the ‘neural flicker’ with classification	32
3.3.4	Behavioural data	33
3.3.5	Classification of the original EEG data without ICA	34
3.3.6	Event-related potentials ‘left’ versus ‘right’	34
3.3.7	Critical frequencies 67 Hz and 77 Hz	34
3.4	Discussion	35
3.4.1	Effects of the shutter glasses on the visual cortex and the human perception	35
3.4.2	Behavioural flicker fusion threshold	37
3.4.3	Why ICA is essential for this analysis	38
3.4.4	Assumption and limitation	38
3.4.5	Relation between neurally-detectable flicker and long-term user satisfaction	38
3.4.6	Conclusion	39
3.5	Lessons learned	39
4	Real-time inference of word relevance from EEG and eye gaze	41
4.1	Introduction	41
4.2	Materials and methods	43
4.2.1	Calibration	43
4.2.2	Online prediction	44
4.2.3	Experimental setup	45
4.2.4	Data acquisition	47
4.3	Results	47
4.3.1	Calibration	47
4.3.2	Online prediction	49
4.4	Discussion	51
4.4.1	Calibration	51
4.4.2	Online prediction	52
4.4.3	Conclusion	53
4.5	Lessons learned	54
5	Variable salience challenges the inference from EEG and eye gaze	55
5.1	Introduction	55
5.2	Materials and methods	57
5.2.1	Experimental design	57
5.2.2	Experimental setup	59
5.2.3	Data acquisition	59
5.2.4	Data analysis	60
5.3	Results	64
5.3.1	Compliance check	64
5.3.2	Target estimation with EEG and eye tracking features	64

5.3.3	Characteristics of target and distractor EEG epochs	66
5.3.4	Eye gaze characteristics	67
5.4	Discussion	70
5.4.1	Compliance check	70
5.4.2	Target estimation with EEG and eye tracking features	70
5.4.3	Characteristics of target and distractor EEG epochs	73
5.4.4	Eye gaze characteristics	75
5.4.5	Interference of eye movements with the EEG	75
5.4.6	Limitations	77
5.5	Conclusion	78
5.6	Lessons learned	79
6	Generalisation properties of the BCI-based relevance detector	81
6.1	Introduction	81
6.2	Materials and methods	82
6.2.1	Experimental design	82
6.2.2	Experimental setup	83
6.2.3	Data acquisition	83
6.2.4	Data analysis	84
6.3	Results	86
6.3.1	Single-trial classification	86
6.3.2	Spatio-temporal dynamics	87
6.3.3	Behavioural performance	87
6.4	Discussion	88
6.4.1	Single-trial classification	88
6.4.2	Spatio-temporal dynamics	89
6.4.3	Behavioural performance	90
6.4.4	Limitations	90
6.5	Conclusion	90
6.6	Lessons learned	91
7	Discussion	95
7.1	Summary	95
7.2	Limitations	96
7.3	Outlook	97
	Bibliography	123

List of Figures

2.1	BCI loop	10
2.2	Exemplary stimulus sequence in an ERP-based BCI	11
2.3	Cortical sources and EEG electrodes	14
2.4	Extraction of spatio-temporal features	17
2.5	Linear classifier	19
3.1	Experimental design	28
3.2	Experimental setup	29
3.3	EEG epochs of the two classes	31
3.4	Independent component selected as neural flicker correlate	33
3.5	Classification results for the single participants	34
3.6	Average classification results and detection rates	35
3.7	Classification results for the original EEG data without ICA	36
3.8	Correlation coefficients of EEG epochs and class labels	36
3.9	Individual classification results for the critical frequencies	37
4.1	Words and semantic categories	42
4.2	Experimental setup	46
4.3	EEG patterns during the calibration phase	48
4.4	Evolution of the scores	49
4.5	Evolution of the rank	50
4.6	EEG patterns during the online phase	52
5.1	From stimulus sequences to an eye tracker guided EEG analysis	55
5.2	Items with variable salience of discriminative information	56
5.3	Distractor, foveal and peripheral target	57
5.4	Illustration of the gaze contingent search task	58
5.5	Experimental setup	60
5.6	ERPs aligned to appearance and fixation of targets and distractors	67
5.7	Statistical differences between target and distractor EEG epochs	68
5.8	EEG classification with spatial features	69
5.9	EEG classification with temporal features	70
5.10	Fixation durations	71
5.11	Latencies between first appearance and first fixation	71

List of Figures

5.12 EEG epochs for short/long fixation durations	76
6.1 Exemplary stimulus sequence and tasks	83
6.2 Classification results within-participants	87
6.3 Classification results across-participants	88
6.4 Effect of the amount of training data on the classification across-participants .	89
6.5 EEG time courses at the midline electrodes	92
6.6 Spatio-temporal EEG activity	93
6.7 EEG activity as scalp topographies	94

List of Tables

3.1	The ten selected shutter frequencies	32
4.1	Final ranks	51
5.1	Classification results for the modalities and conditions	65
5.2	Results of the combined classifier and the split analysis	66
5.3	Fixation frequencies	69
5.4	Average duration of the saccades	72

1 Introduction

1.1 Overview

The brain is the centre of the nervous system, integrates the sensory impressions, gives rise to the conscious perception of the world, plans actions, and sends commands to the muscles. Technical applications and devices augment the human capacities for perception and action, and are usually designed with the user in view for a smooth and efficient usage. Due to the fundamental role of the brain for perception and action, also the mind of the user should be taken into consideration – and not only the physical ergonomics. Straightforward access to the mind is possible by introspection, subjective reports, user questionnaires, behavioural observations, and by allowing for interaction with a device (e.g. with a computer via mouse and keyboard). Yet, the most immediate access to the mind can arguably be achieved by directly observing the neural processes that occur in the brain.

Through brain-computer interfacing (BCI), specific brain activity patterns, which correspond to the mental processes of a person, can be recognised in the electroencephalogram (EEG) in real-time. The detection of these patterns makes it possible to send volitional commands to computers by the power of thought alone – without relying on any muscle movements. BCI methods can potentially also open a novel ‘window on the mind’ of device users by detecting user-related information hidden in the brain signals. Previous investigations have already expanded the scope of BCI beyond volitional control and communication. It was shown that BCI methods can serve for improving the ergonomics of devices and software, and it was suggested that implicit information decoded online from brain signals can support vehicle and machine operators (cf. section 2.2 for details and references).

In this dissertation, the benefits of inferring user-related information from non-invasive recordings of brain activity were demonstrated in two lines of approach. First, it was demonstrated that a usability flaw of a technical device can be uncovered with multivariate data analysis methods employed in BCI. Test persons were not able to notice the deficiency under investigation due to the limits of human perception. The device imposed an imperceptible neural workload on the brain, which is potentially uncomfortable on the long-term. The

findings suggest a possible remedy, which may help to improve the ease of use of the assessed device. In a second line of research, it was shown that the subjective relevance of viewed items can be estimated based on EEG and eye tracking signals. The resulting relevance map of the visual surrounding makes it possible to infer the current interest of the individual user in real-time, which could be taken into account by novel types of adaptive, personalised software.

1.2 Outline of the dissertation

- In chapter 2, fundamentals and previous work related to the decoding of brain signals are discussed. It is substantiated that neural processes in the human brain are complex and can be observed only to a very limited extent. First, a brief primer on the human brain is given, and it is detailed how neural activity can be measured non-invasively. Second, previous investigations on brain-computer interfacing and mental state decoding are introduced, which trace signatures of mental processes in brain signals in real-time. Third, it is explained how information hidden in the signals can be uncovered with methods from machine learning, which can recognise multivariate patterns that are obscured by irrelevant activity and noise.
- In chapter 3, it is shown that BCI methods can provide an objective measure of the workload that a technical device demands from the brain of the user. Crucially, the BCI-based approach proved to be particularly sensitive and could uncover a usability flaw that test persons did not notice, due to the limits of human perception. Traces of an imperceptible (but avoidable) neural processing effort were detected in the brain signals. Discovering such negative effects can turn out to be critical for the sustained success of a product. Users may decide against a product, being unaware of the exact reason, when they are exposed to an imperceptible neural strain on the long-term. Specifically, the neural processing effort imposed by shutter glasses on the viewer of stereoscopic television was quantified with electroencephalography as a function of the shutter frequency. At low shutter frequencies, an annoying and fatiguing flicker is perceived, which diminishes with increasing frequency and vanishes above a critical frequency. For an optimal viewing comfort, the shutter glasses run at a frequency above the perception threshold. Nevertheless, effects of the shutter glasses on the brain were detected also for common shutter frequencies, and up to about 20 Hz above the flicker perception threshold – but vanished at higher frequencies. Increasing the shutter frequency accordingly can potentially prevent visual fatigue, by avoiding the unnecessary neural workload.
- From chapter 4 onwards, it is demonstrated that the subjective relevance of the visual surrounding can be mapped with information implicitly contained in EEG and eye tracking signals. Specifically, it was demonstrated that it is possible to decode online from the recorded signals which words were subjectively relevant for a reader. The marking of relevant words in a text with a highlighter pen could be an analogy for conveying the idea. Indicating which items in the visual surrounding are subjectively

relevant is time-consuming and distracting, and can be distorted by a response bias, which can potentially be circumvented with the BCI-based approach. The obtained implicit relevance information can be aggregated in a user interest profile, that can be updated in real-time. Novel types of adaptive software could take this information about the user interest into account. In addition, time-resolved relevance maps of the field of view can help to improve the usability of devices, websites, or stores.

- In chapter 5, a problem related to the inference of relevance maps from EEG and eye movements was tackled. In typical BCI experiments, sequences of single stimuli are flashed. The stimulus onset serves as reference time point for the extraction of feature vectors from the continuously recorded EEG. Thus, it can be assumed that the informative neural activity is tightly time locked to the reference time point. In contrast, time locked activity can not be taken for granted when *several* items are present in the field of view at the same time, and not only a *single* stimulus. The eye gaze jumps with saccades from one fixated position to the next. Thus, the saccades can serve as reference time points for the feature extraction. However, the regular visual surrounding contains diverse items of a variable salience, ranging from a subtle modesty to an eye-catching flamboyance. This salience spectrum can be problematic for the estimation of relevance maps for the following reason. Light entering the eye along the line of sight falls onto the fovea where the retina provides a better resolution than peripheral retinal areas. Therefore, an item of low salience may be recognised only when the centre of gaze has ‘landed’ on the item, which is now captured in high-resolution by the fovea. In contrast, a highly salient item can be recognised already in peripheral vision, i.e. before or even without a saccade towards the item. For this reason, the timing of recognition with respect to the saccades (used as reference time points) varies with the salience. Accordingly, it was tested if this temporal jitter prevents the estimation of relevance maps from EEG and eye gaze.
- In chapter 6, the generalisation properties of the BCI-based relevance detector were inspected. ‘Relevance’ has to be artificially created in experiments, and an intrinsic interest has to be mimicked with some mental task. For instance, the task could be given to count all relevant items. Focusing the attention in this way is legitimate for a BCI application for communication. However, the intended *implicit* relevance detection would not be possible if the BCI merely detects effects of the specific mental task in the brain signals. Accordingly, it was evaluated if the neural activity detected by the BCI is task specific, or – alternatively – if the detector can generalise over different scenarios. Good generalisation properties might indicate that the subjective experience of considering something as relevant is captured indeed, and not merely the performance of a specific, artificial task.
- In chapter 7, the insights gained are summarised in a concluding discussion, limitations are stated that have to be considered, and an outlook to possible future research is given.

1.3 List of included publications

The chapters 3, 4, 5, and 6 are based on the following publications in this order.

Wenzel, M. A., Schultze-Kraft, R., Meinecke, F. C., Cardinaux, F., Kemp, T., Müller, K.-R., Curio, G., and Blankertz, B. (2016d). EEG-based usability assessment of 3D shutter glasses. *Journal of Neural Engineering*, 13(1):016003. doi: 10.1088/1741-2560/13/1/016003¹

© IOP Publishing. Reproduced with permission.

Wenzel, M. A., Bogojeski, M., and Blankertz, B. (2017). Real-time inference of word relevance from electroencephalogram and eye gaze. *Journal of Neural Engineering*. doi: 10.1088/1741-2552/aa7590²

© IOP Publishing. Reproduced with permission.

Wenzel, M. A., Golenia, J.-E., and Blankertz, B. (2016c). Classification of eye fixation related potentials for variable stimulus saliency. *Frontiers in Neuroprosthetics*, 10(23). doi: 10.3389/fnins.2016.00023³

Wenzel, M. A., Almeida, I., and Blankertz, B. (2016b). Is neural activity detected by ERP-based brain-computer interfaces task specific? *PLoS ONE*, 11(10):1–16. doi: 10.1371/journal.pone.0165556⁴

1.4 Additional publications

Contributions to the following articles are related to the topic of this dissertation:

Golenia*, J.-E., Wenzel*, M. A., Bogojeski, M., and Blankertz, B. (2017). Implicit relevance feedback from electroencephalography and eye tracking in image search. *Submitted*. *) Equal contribution

Blankertz, B., Acqualagna, L., Dähne, S., Haufe, S., Schultze-Kraft, M., Sturm, I., Ušćumlić, M., Wenzel, M. A., Curio, G., and Müller, K.-R. (2016). The Berlin Brain-Computer Interface: Progress beyond communication and control. *Frontiers in Neuroscience*, 10:530. doi: 10.3389/fnins.2016.00530

¹Concept by MW et al. Implementation, data acquisition and analysis by MW. Manuscript drafted by MW and revised by RSK, KM, GC and BB.

²Concept by MW and BB. Implementation and data acquisition by MW, MB and BB. Data analysis by MW and MB. Manuscript drafted by MW and revised by MB and BB.

³Concept by MW and BB. Implementation and data acquisition by MW and JG. Data analysis by MW. Manuscript drafted by MW and revised by JG and BB.

⁴Concept by MW and BB. Implementation and data acquisition by MW, IA and BB. Data analysis by MW. Manuscript drafted by MW and revised by IA and BB.

1.5 Conference contributions

Related contributions to conferences and workshops:

Wenzel, M. A., Almeida, I., and Blankertz, B. (2016a). The contribution of counting to neural activity evoked by the oddball paradigm. In *Proceedings of the 6th International Brain-Computer Interface Meeting*, Asilomar, USA

Wenzel, M. A., Moreira, C., Lungu, I.-A., Bogojeski, M., and Blankertz, B. (2015). Neural responses to abstract and linguistic stimuli with variable recognition latency. In Blankertz, B., Jacucci, G., Gamberini, L., Spagnoli, A., and Freeman, J., editors, *Symbiotic Interaction*, volume 9359 of *Lecture Notes in Computer Science*, pages 172–178. Springer International Publishing. doi: 10.1007/978-3-319-24917-9_19

Golenia, J.-E., Wenzel, M. A., and Blankertz, B. (2015). Live demonstrator of EEG and eye-tracking input for disambiguation of image search results. In Blankertz, B., Jacucci, G., Gamberini, L., Spagnoli, A., and Freeman, J., editors, *Symbiotic Interaction*, volume 9359 of *Lecture Notes in Computer Science*, pages 81–86. Springer International Publishing. doi: 10.1007/978-3-319-24917-9_8

Arndt, S., Wenzel, M. A., Antons, J.-N., Köster, F., Möller, S., and Curio, G. (2014). A next step towards measuring perceived quality of speech through physiology. In *Proceedings of INTERSPEECH 2014*, pages 1998–2001, Singapore

Schultze-Kraft, R., Gorgen, K., Wenzel, M. A., Haynes, J. D., and Blankertz, B. (2013). Cooperating brains: Joint control of a dual-BCI. In *Proceedings of the 5th International Brain-Computer Interface Meeting*, Asilomar, USA

2 Fundamentals

In this chapter, it is explained how neural processes in the brain can be measured, previous research on brain-computer interfacing is introduced, and it is specified how information hidden in brain signals can be uncovered with multivariate methods from machine learning and signal processing.

2.1 Measuring neural processes in the human brain

The human brain serves as control centre of the body, receives information from the sensory organs and sends commands to the muscles via nerves. It comprises a network of about 86 ± 8 billion neurons, supported by about the same number of non-neuronal cells [Azevedo et al., 2009]. A typical neuron in the brain is a cell that branches into an axon and multiple dendrites in order to connect via synapses with a large number of other neurons. Each neuron receives information from approximately ten-thousand neurons, in some cases even from hundred-thousand neurons, over synapses at its dendrites and cell body, and transmits information to roughly thousand neurons, over synapses at its axon [Kandel et al., 2000, Part III, V and VI].

Within this network of neurons, information is transmitted and processed electro-chemically [Kandel et al., 2000, Part II and III]. A typical neuron transmits information to a connected neuron by releasing neurotransmitters into the synaptic cleft. The neurotransmitters float towards the membrane of the connected (= postsynaptic) neuron and bind to receptors that influence the permeability of ion channels in the cell membrane. As a consequence, postsynaptic transmembrane currents and thus the voltage between interior and exterior of the neuron temporary change (the electric potential inside and outside of a neuron differs 'at rest', due to proteins that pump ions through the cell membrane). The temporary voltage changes originating at the single synapses spread over the cell membrane of the neuron, which integrates the information received from all connected neurons [Koch and Segev, 1998]. If the summary voltage at the origin of the axon surpasses a threshold, an action potential is elicited and is actively transmitted avalanche-like along the axon, which leads to an emission of neurotransmitters at the synapses, that again influence other neurons [Kandel et al., 2000;

Chapter 2. Fundamentals

Hallez et al., 2007].

Very complex computations and learning are possible as a result of the intricate wiring of this large network of connected neurons and of the plasticity of the synaptic connections [Abeles, 1991; Koch and Segev, 1998; Koch, 2004; Trappenberg, 2009].

The neural processes, very briefly outlined here, can be observed at different scales with a wide variety of methods from inside and from outside of the skull. Invasive methods, such as implantable microelectrodes or multielectrode arrays, patch or voltage clamping, electrocorticography, optical imaging with voltage-sensitive dyes, calcium imaging, or optogenetics, measure neural processes at close range but put the health at risk [e.g. Hubel and Wiesel, 1962; Grinvald et al., 1988; Ulbert et al., 2001; Stosiek et al., 2003; Miller et al., 2007; Knöpfel, 2012]. Mass activity of large populations of neurons can be captured from outside of the skull with non-invasive methods such as electroencephalography (EEG) and magnetoencephalography (MEG), or, via related changes in the blood flow, with functional magnetic resonance imaging (fMRI) and functional near-infrared spectroscopy (fNIRS). EEG offers some advantages in comparison to other non-invasive methods, because it does not require huge, immobile and very expensive instruments and shielded rooms like MEG and fMRI, and because it provides a higher temporal resolution than fNIRS (and fMRI) [Cohen et al., 1968; Cohen, 1972; Hämäläinen et al., 1993; Logothetis, 2008; Ferrari and Quaresima, 2012].

Work on the EEG dates back to the nineteenth century with animal experiments of Richard Caton and Adolph Beck [Caton, 1875; Beck, 1888, 1890]. Half a century later, Hans Berger recorded the human EEG for the first time [Berger, 1929; Brazier, 1959, 1961]. For the EEG recording, several electrodes are placed on the scalp, in a grid like arrangement, and electric potential differences between electrodes are measured, amplified, (and nowadays) digitised, and sent to a computer, where they are saved and processed. Until recently, EEG systems had been inconvenient to use, because the conductivity between electrodes and scalp had to be improved in a lengthy procedure with conductive gel, that had to be washed out later. Besides, the bulky and cable-bound equipment had to remain stationary. However, usability and mobility of the systems are being improved at fast pace with recent technological innovations like mobile EEG systems [Stopczynski et al., 2014; De Vos et al., 2014; Mullen et al., 2015] with gel-free [Popescu et al., 2007; Grozea et al., 2011; Zander et al., 2011; Guger et al., 2012], miniaturized [Nikulin et al., 2010] electrodes that can be placed hardly visible in/on/around the ear [Looney et al., 2014; Debener et al., 2015; Norton et al., 2015; Goverdovsky et al., 2016a,b]. Moreover, large technology companies and customers are being interested more and more in wearable physiological sensors [Piwek et al., 2016]. In-ear headphones with different physiological sensors including EEG, which connect with a smartphone, are under development (e.g., 'The Aware' from 'United Sciences', Atlanta, USA). Deployable systems will make EEG measurement during daily life more and more feasible, even though the signal quality may remain limited (as well as the potential number and locations of the sensors on the head).

2.2. Brain-computer interfacing and mental state decoding

The measured EEG signal is generated by electric currents through the membranes of neurons and other cells (cf. above in this section). The transmembrane currents dynamically change the electric potential of the extracellular medium at each location within the brain. Synchronous activity of a large number of adjacent and parallel oriented cells results in (primary and secondary) currents that do not mutually cancel out but that are transmitted via volume conduction through the tissue, and that can result in measurable electric potential differences on the scalp. Postsynaptic transmembrane currents and related return currents dominate over other sources such as action potentials in neurons or slow fluctuations in non-neuronal cells. The most important contributors to the EEG signal are presumably postsynaptic transmembrane currents at the parallel oriented apical dendrites of the numerous pyramidal cells in the cortex [Telenczuk et al., 2011; Buzsáki et al., 2012; Einevoll et al., 2013].

The large resistivity of skull and skin, in comparison to brain tissue and cerebrospinal fluid, limits the possible spatial resolution of EEG [Malmivuo and Suihko, 2004]. Moreover, the measured signal is disturbed by artefacts, because electric activity of muscles and movements of the eyes, which constitute dipoles, result in large amplitude signals that interfere with the neural signals [Urigüen and Garcia-Zapirain, 2015]. Besides, every source (e.g., a cortical column receiving synchronous input) can affect the signal measured at different positions on the scalp, due to volume conduction. In addition, many single sources can be active at different locations within the brain at the same time and sum up to the signal measured on the scalp [Parra et al., 2005]. Disentangling and localising the single sources is difficult because numerous combinations of an unknown number of presently active sources can potentially generate the signal on the scalp (cf. also the discussion of the linear generative model of the EEG in section 2.3). The problem of source localisation can be tackled by modelling the transmission of the signal from source to sensor through the biological tissue (forward modelling) and by inferring the most probable source activity given the measured data (backward modelling) [e.g., Darvas et al., 2004; Hallez et al., 2007; Grech et al., 2008].

In summary, non-invasive observation of the complex neural processes is possible only to a rather limited extent. Accordingly, suitable experimental paradigms and appropriate data analysis methods are required in order to make inferences about neural processes in the brain in a meaningful way.

2.2 Brain-computer interfacing and mental state decoding

Every thought or mental process is presumably represented by the electro-chemical processes in the brain. However, only a very limited view on these processes is possible with the available measurement technology, as it was argued in the previous section 2.1. Research on brain-computer interfacing and mental state decoding has taken up this challenge in order to detect traces of the mental processes in the sensor data in real-time.

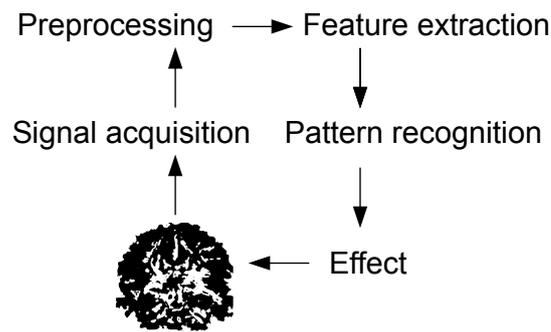


Figure 2.1: The user sends commands to the computer by the power of thought, i.e. by volitionally generating neural ‘signatures’. The brain-computer interface recognises neural activity patterns, and translates the estimated intention of the user into a noticeable effect, such as the selection of a letter or the movement of a robot arm. For pattern recognition, brain signals are acquired and preprocessed, and feature vectors are extracted from the continuous signals.

2.2.1 Brain-computer interfacing for communication and control

Brain-computer interfacing typically aims at providing a command signal under volitional control that is only based on signals originating in the brain. Exemplary use cases are cursor and prosthesis control, communication via speller applications, and switches that allow for choosing from different options, most notably in applications intended for paralysed users [Millán et al., 2010]. BCIs make control possible by the power of thought alone, usually without involving any muscle movements or peripheral nerves (note that the ‘I’ in the abbreviation ‘BCI’ can stand both for ‘interfacing’ and ‘interface’, depending on the context). Extensive research has established a (limited) set of experimental procedures that allow humans to intentionally produce neural ‘signatures’ that can be detected in the measured signals, and translated to an effect (cf. figure 2.1). A brief overview is given in the following.

Humans (and monkeys) can learn to control the firing pattern of single neurons or populations of neurons. The patterns can be detected and translated to control signals that make it possible to steer devices such as robot arms [e.g. Schmidt, 1980; Wessberg et al., 2000; Nicolelis, 2001; Musallam et al., 2004; Carmena et al., 2003; Hochberg et al., 2006; Schwartz et al., 2006; Velliste et al., 2008; Collinger et al., 2013]. Invasive measurement techniques are required in this case and offer a comparably reliable control. Non-invasive methods such as EEG (cf. section 2.1) have the advantage of presenting less risk to health. People can learn to volitionally regulate slow cortical potentials measured with EEG [e.g. Elbert et al., 1980; Birbaumer et al., 1999], or can learn to modulate the power of EEG oscillations in certain frequency bands by imagining limb movements [e.g. Wolpaw et al., 1991; Kübler et al., 2005; Pfurtscheller et al., 2006; Höhne et al., 2014; Vansteensel et al., 2016]. The modulation, detected by the BCI, can be used as control signal. Learning is made possible or supported by biofeedback [cf. Sitaram et al., 2017, and also figure 2.1]. Time and effort can be saved if the burden to learn is partly transferred to the computer [Blankertz et al., 2002, 2007, 2008].

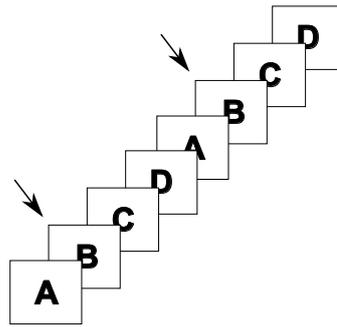


Figure 2.2: Exemplary stimulus sequence presented by a BCI (that is based on event-related potentials). The options A, B, C and D are rapidly flashed on the screen in a repeated sequence. The user can select an option by focussing on the corresponding stimulus (e.g. by silently counting how often the stimulus appeared, which makes it easier to direct the attention). The BCI infers the selection of the user (here: B, indicated by the arrows) from spatio-temporal patterns present in the EEG [inspired by the ‘centre speller’ in Treder et al., 2011, where first a group of letters and then a single letter can be chosen].

The BCI paradigm based on event-related potentials (‘ERPs’) exploits the phenomenon that attention modulates the neural response to stimuli [e.g. Farwell and Donchin, 1988; Treder et al., 2011; Acqualagna and Blankertz, 2013]. This paradigm will be of particular importance in later chapters of this dissertation. Stimuli of interest can be discriminated from other stimuli based on spatio-temporal patterns in the EEG [e.g. Sutton et al., 1965; Picton, 1992; Polich, 2007]. The phenomenon applies to the different sensory domains of sight, hearing and touch. Therefore not only visual, but also auditory or tactile stimuli can be employed, which is crucial when the visual capacities are limited [Brouwer and Van Erp, 2010; Belitski et al., 2011; Schreuder et al., 2011; van der Waal et al., 2012; Höhne and Tangermann, 2014]. Control is provided by the ERP-based BCI as follows. Different stimuli represent different options, e.g. letters or commands, and are presented in a sequence. The option of interest can be selected by directing the attention towards the corresponding stimulus (cf. figure 2.2). Learning of the human is not required. Instead, the computer learns to recognise spatio-temporal patterns in the EEG based on labelled training data recorded during a calibration session. For calibration, the user is told to focus on certain stimuli while ignoring other stimuli [Blankertz et al., 2011]. In the subsequent online application phase, the mentally selected option can be inferred from the signals with the previously trained classification function. The repeated presentation of the stimuli (e.g. in a randomised loop) allows for accumulating evidence and estimating the selected option with greater confidence. The attention is typically directed by silently counting how often the selected stimulus appeared among the other stimuli (cf. chapter 6).

Besides, neural activity patterns corresponding to different mental tasks [Millán et al., 2002; Millán and Mouriño, 2003], and to making mistakes (‘error potential’) [Schalk et al., 2000; Schmidt et al., 2012] can be detected in the EEG. Other approaches exploit ‘steady state’ or ‘code modulated’ visually or auditory evoked potentials that can be regulated by selective attention [Sutter, 1984; Cheng et al., 2002; Farquhar et al., 2008; Bin et al., 2011; Hwang et al.,

2012; Waytowich and Krusienski, 2015; Thielen et al., 2015; Tsoneva et al., 2015]. Also other non-invasive measurement modalities such as near-infrared spectroscopy can be used instead of – or in addition to – electrophysiology [e.g. Sitaram et al., 2007; Fazli et al., 2012; von Lüthmann et al., 2015, 2016].

A BCI system typically comprises a feedback loop of different components for (1) the acquisition of the signals from the brain of the user, (2) preprocessing of the signals, (3) feature extraction from the signals, (4) pattern recognition in order to estimate the intention of the user, which is (5) translated to an effect that can be noticed by the user, e.g. the selection of a letter or the movement of a robot arm [cf. figure 2.1 and van Gerven et al., 2009; Nicolas-Alonso and Gomez-Gil, 2012, for a detailed treatise]. Roots of the investigations on brain-computer interfacing can be traced back to the 1960s and 1970s [e.g. Sutton et al., 1965; Donchin, 1966; Vidal, 1973, 1977; cf. also the introduction of Blankertz et al., 2016]. Overviews of the field of research are given, e.g., in Wolpaw et al., 2002; Lebedev and Nicolelis, 2006; Dornhege et al., 2007; Mak and Wolpaw, 2009; Wolpaw and Wolpaw, 2012 and Hwang et al., 2013.

2.2.2 Mental state decoding

EEG served for a long time as medical device in the clinic for monitoring sleep stages and epileptic seizures [e.g. Campbell, 2009; Acharya et al., 2013]. Recently, EEG-based applications for mental state monitoring gained increasing attention (cf. section 3 in Blankertz et al., 2010 and Müller et al., 2008; Zander and Kothe, 2011; Blankertz et al., 2016), that could expand the scope of BCI beyond the provision of a communication channel for the paralysed (cf. section 2.2.1). Mental states, e.g. related to alertness, arousal, attention, fatigue, mental workload, task engagement, or to the preparation of movements, could be detected in real-time, which is potentially useful for a wide range of purposes [Pope et al., 1995; Makeig et al., 1996b; Gevins and Smith, 2003; Berka et al., 2007; Shen et al., 2008; Müller et al., 2008; Stikic et al., 2011; Ayaz et al., 2012; Baldwin and Penaranda, 2012; Brouwer et al., 2012; Harrivel et al., 2013; Treder et al., 2014; Mühl et al., 2014; Schultze-Kraft et al., 2016a]. Not only EEG, but also fMRI [Haynes and Rees, 2006] and fNIRS [Ayaz et al., 2013] signals can be decoded in order to continuously track covert mental states.

Brain state monitoring could support drivers, pilots, air traffic controllers, machine operators, or computer users [Kohlmorgen et al., 2007; Müller et al., 2008; Haufe et al., 2011; Aricò et al., 2016; Borghini et al., 2014; Schultze-Kraft et al., 2016b]. Human errors could be avoided by issuing warnings when the person is inattentive in a safety critical moment, or by offering assistance when the person is overwhelmed with the present situation. Moreover, the design of cockpits and machines might be objectively quantified with respect to the ease of use, during task execution and without the subjective bias of time-consuming questionnaires [Müller et al., 2008], which can be subsumed under the term ‘neuroergonomics’ [Parasuraman, 2003; Parasuraman and Rizzo, 2008; Parasuraman et al., 2012; Mehta and Parasuraman, 2013].

Certain usability flaws can not be detected with questionnaires, when the limits of human

2.3. Multivariate methods for detecting information in the EEG

perception are crossed, but can potentially be revealed with EEG-based BCI methods [e.g. Porbadnigk et al., 2011, in the case of lighting]. Imperceptible effects can turn out to be critical for the success of a product, when the users are exposed to negative effects on the long-term. In this case, the users may decide against the product, being unaware of the exact reason. Not only the quality of hardware can be evaluated with EEG, but also the quality of software [e.g. codecs for audio and video compression: Porbadnigk et al., 2010, 2013; Antons et al., 2010, 2012; Creusere et al., 2012; Scholler et al., 2012; Arndt et al., 2012; Mehta and Kliewer, 2015; Acqualagna et al., 2015], or of stereoscopic visualisations [Frey et al., 2016], by directly measuring neural processes related to perception.

Mental state detectors may facilitate the interaction between humans and machines, which is unbalanced at present [Hettinger et al., 2003]. Devices can communicate rich information about their current state to the operator. In contrast, the operator can only send limited information to the machine via conventional controllers such as keyboard, mouse, buttons, or sliders. Letting the device learn more about the individual user according to information implicitly contained in the EEG, could balance and improve the interaction (cf. section 4 and Parra et al., 2003; Vidaurre et al., 2011a,b; Pohlmeier et al., 2011; Zander and Kothe, 2011; Ušćumlić et al., 2013; Jangraw et al., 2014; Eugster et al., 2014, 2016; Kauppi et al., 2015; Finke et al., 2016; Blankertz et al., 2016; Müller et al., 2017). Sensor data can be recorded in the background and could augment conventional input devices.

Methods for mental state detection can exhibit a high sensitivity under constrained laboratory conditions, where an isolated parameter is changed. Yet, the sensitivity can stand in contrast to a low specificity to be expected in more realistic situations, where muscle and eye movements interfere and constitute a confounding factor [Brouwer et al., 2015]. For instance, spectral EEG features can be assessed in order to detect band power modulations that correlate positively or negatively with cognitive workload. However, also lively movements can result in broad band signals in the EEG.

The processing steps for mental state decoding are very similar to the steps employed in a BCI for communication and control, but do not necessarily form a feedback loop (cf. figure 2.1), because it is not required that the user learns about the outcome of the prediction.

2.3 Multivariate methods for detecting information in the EEG

Multivariate methods from machine learning and signal processing play a key role for BCI in uncovering information that is hidden in the neural signals. The brain, as underlying signal generator, is characterised by many processes happening at the same time (e.g., perception in the different sensory domains with concurrent planning and execution of movements). Moreover, each process is computed by a large number of cooperating neurons distributed throughout the brain (cf. section 2.1), and different processes can occur in rapid succession. The resulting rich spatio-temporal dynamics of the neural activity, however, is observed only indirectly from outside of the head, at least in the case of EEG. Measurement noise

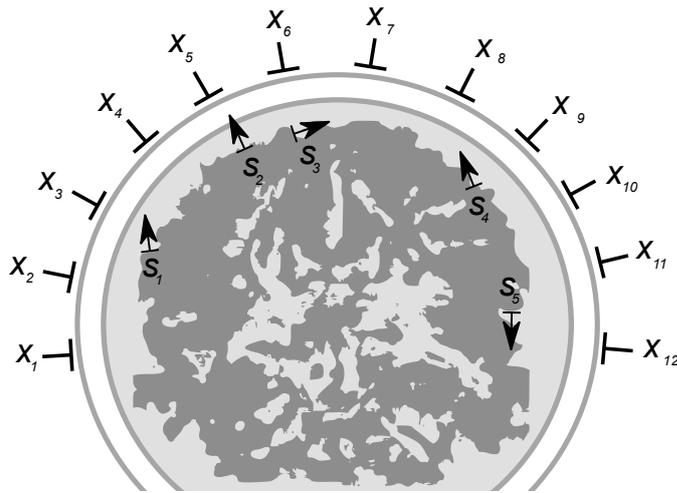


Figure 2.3: Illustration of five dipole sources with different orientations and locations in the cortex. Twelve EEG electrodes on the scalp measure different mixtures of the activity of the five sources. Potentially far more sources can be active within the brain at the same time, in contrast to this example, where the electrodes outnumber the sources. Cerebro-spinal fluid, meninges, skull and scalp cover the brain (represented by layers of different luminance in the illustration).

and artefacts, e.g. due to eye movements, deteriorate the signal quality. In order to capture as much information as possible about the neural processes, signals are recorded at many locations on the scalp in parallel with a high temporal resolution, resulting in high dimensional data. Multivariate methods are particularly suited for making inferences about the neural processes from the high dimensional data, because they can detect patterns that are potentially distributed over different dimensions, and that are obscured by measurement noise and ‘irrelevant’ neural activity from the ‘background’.

2.3.1 Linear generative model of the EEG

Electrical potential differences measured at time t between M electrodes on the scalp and a reference electrode can be represented by a vector $\mathbf{x}(t) = [x_1(t), \dots, x_M(t)]^T$. Due to volume conduction, the potential measured at a single electrode does not only reflect the activity in the cortex directly below the respective electrode position, but a mix of signals from different cortical sources at different locations in the brain, as well as from other non-cortical sources such as eyes or muscles (cf. section 2.1 and the illustration in figure 2.3).

The activity of each source $s(t)$ at time t contributes linearly to the measurement on the scalp $\mathbf{x}(t)$, according to the linear generative model of the EEG [cf. section ‘Linear model for EEG’ in Parra et al., 2005]).

$$\mathbf{x}(t) = \mathbf{a} s(t) \tag{2.1}$$

2.3. Multivariate methods for detecting information in the EEG

The proportional factors in $\mathbf{a} = [a_1, \dots, a_M]^T$ depend on the source distribution and orientation, on the physical properties of the body tissues, and on the electrode positions and contacts. Multiple sources active at time t can be represented by a vector $\mathbf{s}(t) = [s_1(t), \dots, s_K(t)]^T$, with K being the number of sources (that are modelled). The contribution of each source $k \in \{1, \dots, K\}$ to the measurement at each electrode $m \in \{1, \dots, M\}$ is given by element a_{mk} of the mixing matrix $\mathbf{A} \in \mathbb{R}^{M \times K}$, which is also referred to as ‘forward model’ (and results from concatenating the proportional factors \mathbf{a} from all K sources). At the electrodes on the scalp, a linear mixture of the sources can be observed according to

$$\mathbf{x}(t) = \mathbf{A} \mathbf{s}(t) \tag{2.2}$$

The summary potential resulting from remaining sources, which are arbitrarily not included in the model, can be summarised as ‘noise’ term $\mathbf{n}(t)$ (cf. the discussion in [Parra et al., 2005], where it is highlighted that measurement noise has/is a source, too).

$$\mathbf{x}(t) = \mathbf{A} \mathbf{s}(t) + \mathbf{n}(t) \tag{2.3}$$

2.3.2 Demixing the EEG

Activity originating in the *cortex* of the brain is of interest when recording EEG. Thus, demixing the signals measured on the *scalp* would be desirable in order (a) to remove interfering signals from non-cortical sources, and (b) to inspect the activity of each cortical source separately. However, solving equation 2.2 for $\mathbf{s}(t)$ is difficult, because only the mixture $\mathbf{x}(t)$ can be measured, whereas the mixing matrix \mathbf{A} and the number of sources K are unknown (and thus the dimensionality of \mathbf{A} and of $\mathbf{s}(t)$). As mentioned at the end of section 2.1, source localisation methods tackle this problem by modelling the signal transmission in the head (forward modelling) and by estimating the most probable source activity according to the recorded data (backward modelling) [e.g. Darvas et al., 2004; Hallez et al., 2007; Grech et al., 2008].

A different approach to the problem can be taken with data-driven methods. Independent component analysis (ICA) and other blind source separation methods can separate mixed sources, based on statistical properties of the recorded multivariate data alone. The source activity can be reconstructed to some degree, without prior information about the physical properties of the head. Inspecting the characteristics of the obtained signal components makes it possible to focus on the cortical activity of interest, and to drop components that are assumed to reflect non-cortical sources. ICA projects the multichannel EEG time series to a new coordinate system where the component time courses are as independent as possible [Hyvärinen and Oja, 2000; Hyvärinen et al., 2001]. The scalp potentials $\mathbf{x}(t)$ are linearly transformed with a demixing matrix $\mathbf{W} \in \mathbb{R}^{M \times K}$, which maximises the mutual statistical

independence of the resulting components $\hat{\mathbf{s}}(t)$.

$$\hat{\mathbf{s}}(t) = \mathbf{W}^T \mathbf{x}(t) \tag{2.4}$$

The demixing matrix \mathbf{W} can be computed with different algorithms and criteria for statistical independence [Hyvärinen and Oja, 2000; Parra et al., 2005; Winkler, 2015, which also give a good overview on ICA and its application to EEG or MEG data]. ICA methods can consider (a) higher order statistics such as skewness or kurtosis and/or (b) the temporal structure of the signals [cf. section 2.2.2 in Winkler, 2015, for a comparison of (a) and (b)]. The ICA method *TDSEP*, which stands for *temporal decorrelation source separation*, decorrelates the component time courses *over time* and is particularly suited for demixing EEG signals due to the large temporal autocorrelation of the EEG [Ziehe and Müller, 1998; Ziehe et al., 2000]. The demixing matrix \mathbf{W} is estimated by diagonalising ‘several time-delayed second-order correlation matrices’, simultaneously and in approximation [Ziehe and Müller, 1998]. Non-Gaussianity of the sources is not assumed, unlike ICA methods exploiting higher order statistics [Hyvärinen and Oja, 2000].

Recommended preprocessing steps for ICA are highpass-filtering, centering (here it is assumed that the data have zero mean), and whitening, i.e. decorrelation of the dimensions and ensuring unity variance [cf. Hyvärinen and Oja, 2000; Winkler, 2015]. Assumptions made by ICA are independence and stationarity of the sources, instantaneous linear mixing without delay, and the condition that the number of available sensors equals the number of active sources [Makeig et al., 1996a; Winkler, 2015]. Especially the equality condition for the number of sources and sensors can be problematic [Makeig et al., 1996a]. Even if the assumptions are not fully met, ICA can separate the mixture to some degree. In any case, ICA can not extract more sources than sensors are available. ICA can be applied not only in the temporal domain to signals that are resolved in time and space, like here, but also in the spatial domain, which can be especially interesting for neurophysiological recordings with a high spatial resolution such as fMRI data [Petersen et al., 2000]. In contrast to *independent* component analysis, *principal* component analysis merely decorrelates the component time courses, using orthogonal basis vectors, determined with an eigenvalue decomposition of the covariance matrix of the data.

2.3.3 Recognising patterns in the EEG

Volitional or unconscious modulations of neural activity can be reflected as patterns in the EEG. Recognising these patterns with methods from machine learning [Bishop, 2007; Duda et al., 2012] is essential for BCI and mental state decoding [Blankertz et al., 2002; Lotte et al., 2007; Blankertz et al., 2007; Tomioka and Müller, 2010; Blankertz et al., 2011; Lemm et al., 2011]. For instance, in an attention-based BCI speller, letters are flashed in a sequence (cf. figure 2.2 in section 2.2.1). The user silently selects a letter, mentally focuses on the chosen letter in the flashed sequence and ignores other letters. Thereby, the neural responses to the letters are

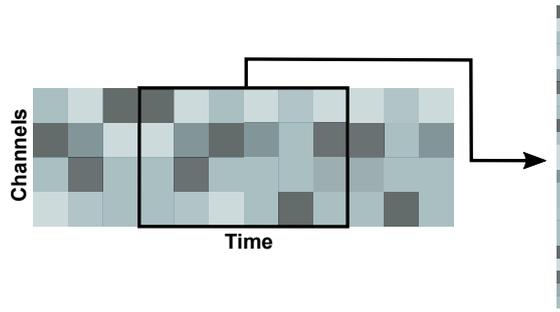


Figure 2.4: A spatio-temporal feature vector is extracted from a short segment of the continuously recorded multichannel EEG. Each square represents the measurement at one time point at one EEG electrode. The measured potentials are illustrated by the luminance of the squares. Actual feature vectors can include more EEG electrodes and time points, and are, therefore, of a higher dimensionality than in this exemplary illustration. The extracted feature vector captures the neural activity in the interval of interest (e.g. the neural response to a stimulus flashed at the beginning of the temporal window).

modulated. The BCI infers the silent selection of the user from corresponding patterns in the EEG.

2.3.3.1 Extracting spatio-temporal features from the EEG

For pattern recognition, feature vectors are extracted from the continuous multichannel EEG time series. In this dissertation, *spatio-temporal* features were inspected that capture the temporal evolution of the EEG potentials measured at several electrode positions on the scalp (cf. illustration in figure 2.4).

Spatio-temporal features can be obtained by cutting out short ($< 1s$) temporal windows from the multichannel EEG following the events of interest (such as the letter ‘flashes’ in the case of a BCI speller). The resulting multichannel EEG epochs indirectly reflect the spatio-temporal dynamics of the underlying neural activity (cf. section 2.3.1). Each EEG epoch can be represented by a feature vector $\mathbf{x} = [x_1, \dots, x_{(M \cdot T)}]^T$, corresponding to M EEG channels and T time points sampled within the chosen temporal window (cf. figure 2.4). A collection of N feature vectors can be gathered in a matrix $\mathbf{X} \in \mathbb{R}^{(M \cdot T) \times N}$.

Optionally, the signals can be preprocessed, for instance by temporal filtering, by spatial filtering (e.g. with ICA; cf. section 2.3.2), by baseline subtraction, and by rejecting epochs contaminated with artefacts – while avoiding certain pitfalls [cf. Table 1 in Lemm et al., 2011].

2.3.3.2 Training a classifier with labelled feature vectors

A classifier can learn to recognise multivariate patterns in the EEG based on training data. Subsequently, the trained classifier can be applied to feature vectors extracted online from the

EEG, or from a separate test set (cf. section 2.3.3.6). Supervised classifier training requires a collection of feature vectors that are labelled with the classes of the corresponding patterns. Such training data can be collected in an initial calibration session, where the classes are known, e.g. by asking the BCI user to focus on certain stimuli in a stimulus sequence or to imagine a movement of either the left or the right hand. Binary classifications between two patterns are made in most BCI applications. Also the selection among multiple options can be simplified to binary decisions, by estimating for each option if it was *either* selected *or* not (e.g. multiple letters in a ‘mental typewriter’). Feature vectors of two classes can be labelled with $y \in \{-1, +1\}$. All N single labels can be gathered in a label vector \mathbf{y} with N elements. A supervised learning algorithm can learn the relation between features and labels based on the training data set (\mathbf{X}, \mathbf{y}) . The relation can be expressed in a classification function that can label a feature vector \mathbf{x} that was not seen yet, and whose class membership is unknown.

2.3.3.3 Linear discriminant analysis

Classifying the feature vectors with a *linear* function is computationally less expensive than more complex non-linear classifiers, and has the advantage of good generalisation properties [Müller et al., 2003; Lemm et al., 2011]. A linear binary classifier bisects the feature space with a separating hyperplane, located where

$$\mathbf{w}^T \mathbf{x} + b = 0 \tag{2.5}$$

Orientation and location of the separating hyperplane in the feature space of D dimensions are determined by the weight (or normal) vector $\mathbf{w} = [w_1, \dots, w_D]^T$, which is perpendicular to the hyperplane, and the bias b . The classification function determines the class membership $y \in \{-1, +1\}$ of a feature vector \mathbf{x} according to

$$y = \text{sgn}(\mathbf{w}^T \mathbf{x} + b) \tag{2.6}$$

Thus, the classifier checks if \mathbf{x} is located in either one or the other section of the feature space separated by the hyperplane (cf. illustration in figure 2.5). The parameters \mathbf{w} and b of the classification function are tuned based on labelled training data. The intention of the parameter tuning is to label feature vectors, that are not part of the training data and whose classes are not known yet, as correct as possible.

Linear discriminant analysis (LDA) is one possibility to tune the parameters and assumes that the feature vectors of the two classes are drawn from two multivariate Gaussian distributions, with different known means, and a common known covariance matrix [Fisher, 1936; Duda et al., 2012]. Fortunately, LDA can perform well even if the assumptions are not fully met [Duda et al., 2012]. Obviously, means and covariance can merely be estimated, and the true distributions are unknown. Moreover, the neural responses consist of different components whose timing can be variable. Thus, variations around the empirical means can not be

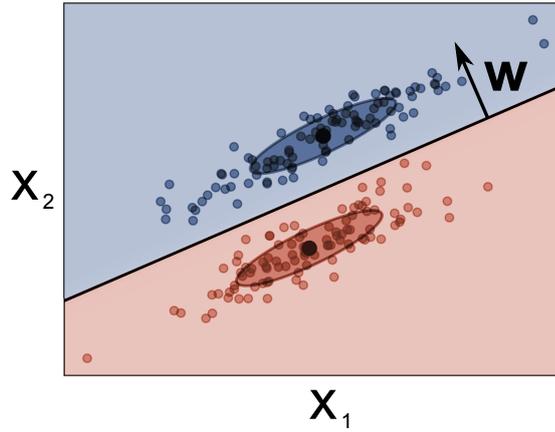


Figure 2.5: Illustration of a linear classifier in two dimensions. Each point represents a feature vector extracted from the EEG (cf. figure 2.4). Blue and red points correspond to two brain states, which can be discriminated here with a linear classification function (with weight vector \mathbf{w}). Note that only the combination of the two dimensions makes the separation possible – the two classes of feature vectors would overlap, when projected on either one or the other axis. In this example, the feature vectors are only two dimensional, in contrast to actual EEG feature vectors that are usually high dimensional. Covariance estimates of the distributions of the two classes are illustrated by ellipses and the mean estimates by larger black dots [code was adapted from Pedregosa et al., 2011, for this illustration].

explained only by a superimposed background activity, modelled by the common covariance matrix. In addition, the distributions can be non-stationary, e.g. when the background activity changes over time. Besides, the quality of the electrode contacts can shift over time. Whether and to what extent the assumptions are fulfilled is discussed in detail in Blankertz et al. [2011].

Based on estimates of the common covariance $\hat{\Sigma}_c$ and of the class-wise means $\hat{\boldsymbol{\mu}}_{+1}$ and $\hat{\boldsymbol{\mu}}_{-1}$, the weight vector \mathbf{w} of the classification function is determined by

$$\mathbf{w} = \hat{\Sigma}_c^{-1} (\hat{\boldsymbol{\mu}}_{+1} - \hat{\boldsymbol{\mu}}_{-1}) \quad (2.7)$$

The dimensionality of the EEG feature vectors is usually large, whereas only few samples are available for training (cf. section 2.3.3.1). In this case, the covariance matrix may be not invertible. As remedy, the pseudoinverse can be applied [Penrose, 1955]. For the estimation of $\hat{\Sigma}_c$ and $\hat{\boldsymbol{\mu}}_{+1}$ and $\hat{\boldsymbol{\mu}}_{-1}$, the training samples of the two classes are split up. The class-wise means $\hat{\boldsymbol{\mu}}_{+1}$ and $\hat{\boldsymbol{\mu}}_{-1}$ are respectively estimated by

$$\hat{\boldsymbol{\mu}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \quad (2.8)$$

The corresponding empirical covariance matrices $\hat{\Sigma}_{+1}$ and $\hat{\Sigma}_{-1}$ are estimated each with

$$\hat{\Sigma} = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{x}_i - \hat{\boldsymbol{\mu}}) (\mathbf{x}_i - \hat{\boldsymbol{\mu}})^T \quad (2.9)$$

which results in the common covariance matrix $\hat{\Sigma}_c$ by averaging

$$\hat{\Sigma}_c = \frac{1}{2} (\hat{\Sigma}_{+1} + \hat{\Sigma}_{-1}) \quad (2.10)$$

The bias b can be chosen in different ways, e.g. as midpoint between the projections of the class-wise means on the normal vector \mathbf{w} of the separating hyperplane.

$$b = -\mathbf{w}^T \frac{(\hat{\boldsymbol{\mu}}_{+1} + \hat{\boldsymbol{\mu}}_{-1})}{2} \quad (2.11)$$

2.3.3.4 Regularisation of the covariance matrix

Feature vectors extracted from the EEG are variable and are characterised by a high dimensionality in comparison to a limited number of available training samples. In this setting, a classification function, that is optimised only with respect to a minimal number of wrong classifications of the training samples may not generalise well to unseen test data. The resulting problem is referred to as *overfitting* to the training data. Linear classifiers can be misled especially by outliers (but are less prone to overfitting due to an inappropriate complexity of the separating hyperplane like non-linear classifiers) [Lemm et al., 2011]. Moreover, the empirical covariance systematically deviates from the true covariance when only a small set of high-dimensional feature vectors is available. In this case, large eigenvalues of the covariance matrix are estimated to be larger than they actually are, and small eigenvalues are believed to be smaller than they are in truth [Blankertz et al., 2011]. As a consequence, the weight vector \mathbf{w} of the classification function is suboptimal (cf. equation 2.7). Regularized linear discriminant analysis can compensate for this systematic deviation by shrinking or stretching the eigenvalues of the covariance matrix towards the average eigenvalue [Friedman, 1989; Tomioka and Müller, 2010].

$$\tilde{\Sigma}(\gamma) := (1 - \gamma) \hat{\Sigma} + \gamma \nu \mathbf{I} \quad (2.12)$$

The shrinkage parameter γ blends between the empirical covariance matrix $\hat{\Sigma} \in \mathbb{R}^{D \times D}$ of the D -dimensional feature vectors, and a spherical covariance matrix characterised by ν , which is the average eigenvalue of $\hat{\Sigma}$. \mathbf{I} is the identity matrix. The optimal shrinkage parameter γ^* can be determined with cross-validation, which is computationally expensive. An alternative is the straightforward computation with an analytic method [Ledoit and Wolf, 2004; Schäfer and Strimmer, 2005], that increases the value of γ^* , the more the covariance estimate varies from

sample to sample [Blankertz et al., 2011].

$$\gamma^* = \frac{N}{(N-1)^2} \frac{\sum_{i,j=1}^D \text{var} \langle (\mathbf{Z}_n)_{ij} \mid n = 1, \dots, N \rangle}{\sum_{i,j=1}^D s_{ij}^2} \quad (2.13)$$

For the computation, the following definitions are made. The covariance matrix of the n -th feature vector \mathbf{x}_n (from the set of N feature vectors; cf. section 2.3.3.1) is defined as

$$\mathbf{Z}_n = (\mathbf{x}_n - \hat{\boldsymbol{\mu}})(\mathbf{x}_n - \hat{\boldsymbol{\mu}})^T \quad (2.14)$$

with $\hat{\boldsymbol{\mu}}$ as average feature vector. The element in row i and column j of the matrix $\hat{\boldsymbol{\Sigma}} - \nu \mathbf{I}$ is denoted by s_{ij} .

2.3.3.5 Feature selection

The high dimensionality of the feature vectors in comparison to the limited number of available training samples has adverse effects on the predictive performance, as discussed in section 2.3.3.4. Regularising the covariance matrix can mitigate this disadvantage. In addition, the dimensionality of the feature vectors can be reduced. For instance, the feature vectors can be projected to a space of a lower dimensionality with principal component analysis, or a subset of particularly informative features can be chosen according to univariate statistical measures (cf. section 2.3.3.7). Besides, the EEG epochs – that build the basis for the spatio-temporal feature vectors – can be downsampled to a lower temporal frequency, or subsampled over specific temporal intervals. In any case, feature selection risks that information is dropped which is potentially valuable for the prediction. Finally, care has to be taken to select features *not* on the basis of *test* data for a correct evaluation of the predictive performance [Lemm et al., 2011, and section 2.3.3.6].

2.3.3.6 Evaluating the classification performance

A classification algorithm learns the relation between features and labels from training data. However, the objective of classifier training is to label feature vectors that have *not been part of the training data* as correctly as possible. Accordingly, it is useful to evaluate how well the generalisation to previously unseen feature vectors is possible (or in other words: to estimate the test error). For instance, the classification performance, that can be expected during the online application of a BCI, can be determined already during the calibration phase. Moreover, the classification performance of different classification algorithms and model parameters can be compared according to independent test samples. Based on this information, the classifier and parameter combination can be chosen that is best suited for the particular problem at hand (cf. the discussion of ‘nested cross-validation’ below).

Information about the training error is less important (‘how well can the classes be separated

in the training set?') in contrast to the test error. Tuning the classifier only with the objective to minimise the training error can lead to overfitting to the training samples and to a weak performance on independent test data. This applies especially to cases with only few but high-dimensional feature vectors available for training. Then, the training samples can be classified perfectly by various classification functions, even if they turn out to be incapable for the classification of test samples.

For the evaluation, the entire data set is split in two complementary partitions. The classification function is trained with the set of labelled (\mathbf{y}) feature vectors (\mathbf{X}) from one partition and applied to feature vectors from the other partition. Then, the estimated labels are compared with the corresponding true labels. The discrepancy between estimated and true labels is referred to as classification *loss* that can be quantified with different metrics, most simply with the 0-1 indicator function, which is 0 if the estimated label and the true label coincide and 1 otherwise. The losses corresponding to all test samples can be summarised as average loss. The true positive rate measures the number of correct classifications of samples from class $y = +1$ in relation to the number of all samples of the class $y = +1$. Vice versa, the false positive rate quantifies how many samples were wrongly labelled as class $y = +1$ in relation to the number of all samples of the class $y = -1$.

Estimated and true labels will coincide with some probability also by random guessing. Therefore, the measured classification performance has to be compared with the chance level that can be expected by random classification. The chance level of the classification loss depends on the class ratio of the samples. In contrast, the area under the curve of the receiver operating characteristic ('AUC ROC') is a metric that is insensitive to class imbalances [Fawcett, 2006]. The chance level of the AUC ROC is 0.5 irrespective of the ratio of the samples from the two classes. Perfect classification is indicated by an AUC ROC of 1. The curve of the receiver operating characteristic is generated by varying the bias in the classification function (cf. equation 2.6), comparing the respective estimated and true labels, and by plotting the true positive rate versus the respective false positive rate.

The classification loss, AUC ROC, or other metric varies with the particular set of samples used for training and testing. This variability can be reduced by repeating training and testing with different partitions of the entire data set, and by averaging the obtained results. This approach is referred to as 'cross-validation' and can be accomplished in different ways [cf. Müller et al., 2001; Lemm et al., 2011], such as

- *n*-fold cross-validation: the entire data set is split into *n* partitions, one respective partition is held out for testing, while all other *n* - 1 partitions are used for training. This procedure is repeated until each partition is tested once (and can be combined with one or more additional shuffles of the samples and corresponding new splits).
- Leave-one-out cross-validation: each sample is tested once, all other respective samples serve for classifier training.

2.3. Multivariate methods for detecting information in the EEG

- Block-wise cross-validation: applicable to experiments where single instances of a class are repeated in succession in a block. All samples per experimental block are always tested together and are never separated. Block-wise cross-validation circumvents the risk to report a *spurious* classification success due to non-stationarities in the EEG, which are unrelated to the class, but which have the effect that the samples are not independent and identically distributed [cf. Lemm et al., 2011].
- Chronological cross-validation: tests are performed only on samples acquired after the training samples (or, at least, on samples from a continuous time period).
- Nested cross-validation: the entire data are split in a training set A and a test set B (with one of the just presented schemes). The training set A is again split in A_a and A_b , which allows for training different models and/or parameter combinations on A_a , choosing the optimal model and/or parameters on A_b , and an independent evaluation on B . The scheme can be repeated with different assignments of the samples to A_a , A_b and B .

2.3.3.7 Characterising the discriminative information

Insights into the underlying reasons for the classification outcome can be gained by characterising the discriminative information. Moreover, the information can be used to select a subset of particularly discriminative features (cf. section 2.3.3.5).

Computing the correlation between each feature and the class labels is a univariate approach to the problem. In the case of spatio-temporal features, the correlation indicates where the univariate discriminative information resides in space and time (“Which electrodes on the scalp and which time points in the EEG epochs are informative?”). The point biserial correlation coefficient r is the measure of choice, because the class labels \mathbf{y} are dichotomous for binary classifications, whereas the features are continuous values [Lev et al., 1949]. The square of the coefficient (r^2) reflects the explained variance. The sign of the original coefficient can be retained, when squaring (signed r^2).

$$r = \frac{\sqrt{\frac{N_{+1}N_{-1}}{N}} (\bar{f}_{+1} - \bar{f}_{-1})}{\left[\sum_{i \in \{+1, -1\}} \sum_{j=1}^{N_i} (f_{ij} - \bar{f}) \right]^{\frac{1}{2}}} \quad (2.15)$$

The collection $\mathbf{X} \in \mathbb{R}^{D \times N}$ contains N feature vectors of D dimensions (cf. section 2.3.3.1). The correlation between each single feature, i.e. each row in \mathbf{X} , and the class labels in \mathbf{y} is computed separately. Observations per feature corresponding to the class $y = +1$ are given by f_{+1j} , $j = 1, \dots, N_{+1}$, and by f_{-1j} , $j = 1, \dots, N_{-1}$ for the class $y = -1$. The number of class members are given by N_{+1} and N_{-1} respectively, and the total number by $N = N_{+1} + N_{-1}$. The mean value of a feature is denoted by \bar{f} and the class-wise means by \bar{f}_{+1} for $y = +1$, and by \bar{f}_{-1} for $y = -1$.

Multivariate classifiers can detect patterns distributed over different features in combination, which goes beyond the just introduced univariate correlation between single features and labels. Interpretation of the multivariate discriminative information requires to transform the classifier weights \mathbf{w} (not interpretable straightforwardly) to a corresponding pattern (interpretable), as detailed in [Haufe et al., 2014].

2.3.3.8 Other possibilities

While the EEG is primarily examined in the time domain in this dissertation, also the frequency domain carries valuable information for BCI. Multivariate statistics can improve the access to this information. An example is the common spatial pattern algorithm [cf. Fukunaga, 1972; Koles et al., 1990; Ramoser et al., 2000; Blankertz et al., 2008] that determines spatial filters which emphasise amplitude modulations in specific frequency bands, e.g. due to imagined limb movements or due to workload changes. The source power comodulation framework and its derivatives decompose the multichannel EEG signal under consideration of the neural oscillations and of internal or external target variables [Dähne et al., 2014a,b; Dähne et al., 2015].

2.4 Lessons learned

- The complex neural processes of the human brain were briefly introduced. Non-invasive observation of the processes is possible only to a very limited extent.
- Brain-computer interfacing and mental state decoding have the objective to detect traces of mental processes in brain signals in real-time.
- Machine learning methods can uncover information hidden in the brain signals, by recognising multivariate patterns obscured by irrelevant activity and noise.

3 EEG-based usability assessment of stereoscopic displays

3.1 Introduction

It is a critical success factor for any product that it can be used easily and with good comfort and, therefore, new devices are extensively tested for their usability during the development process. Neurotechnology can contribute to the usability evaluation by providing objective measures of the demanded neural workload without a potential bias in the subjective judgement. Moreover, usability impediments can be uncovered that are not consciously perceived by test subjects. Previous studies demonstrated that analysis techniques based on the electroencephalogram (EEG) can obtain a higher sensitivity than behavioural measures in the assessment of audio quality [cf. Porbadnig et al., 2010, 2013; Antons et al., 2012]. Similar methods exist in the visual domain [e.g., Mustafa et al., 2012; Scholler et al., 2012; Arndt et al., 2012; Lindemann et al., 2011; Acqualagna et al., 2015]. This relatively new approach draws from the multivariate data analysis techniques that have been developed in brain-computer interface (BCI) research (cf. chapter 2).

The specific problem addressed in this chapter is the neural processing effort imposed on the viewer of 3D television by the shutter glasses. 3D TV creates a spatial impression by presenting two images that are photographed from two slightly different perspectives separately to the two eyes. Active shutter 3D TV is one of several methods to display 3D images and works as follows. While the two stereo channels are shown alternately on the TV screen, the 3D shutter glasses open and close correspondingly to guide each stereo channel only to the respective eye.

Viewers can experience visual discomfort after prolonged watching of stereoscopic 3D content. The vergence-accommodation conflict, crosstalk, flicker, misalignment of the stereo image pair (such as a vertical disparity or a focus mismatch) or unnatural blur were suggested to cause this insufficient ease of use [cf. Bando et al. 2012; Kooi and Toet 2004; Lambooij et al. 2007; Ukai and Howarth 2008; Tam et al. 2011; Woods 2010 for reviews about possible causes for discomfort, and Huynh-Thu et al. 2010; Lambooij et al. 2011a,b about methods to assess 3D quality]. Visual fatigue caused by viewing stereoscopic stimuli was assessed in several

recent physiological studies. The overall effect of 3D TV on visual fatigue was studied with EEG comparing 3D versus 2D [e.g., Kim and Lee, 2011; Ting et al., 2011], before versus after watching [Mun et al., 2012; Chen et al., 2013] and a short versus a long viewing duration [Li et al., 2008]. The effects of the vergence-accommodation-conflict were investigated, which can make stereoscopic images uncomfortable, and it was suggested that “EEG is likely to enable the conception of adaptive systems, which could tune the stereoscopic experience according to each viewer” [Frey et al., 2014]. With functional magnetic resonance imaging, brain activity of subjects watching 3D and 2D films was compared [e.g., Gaebler et al., 2014] and the relationship was investigated between binocular disparity, visual fatigue and the visual system of the brain when watching stereoscopic stimuli [Kim et al., 2011, 2014; Jung et al., 2013]. Furthermore, “viewer discomfort” was quantified “by measuring eye blinking rates” [Cho and Kang, 2012].

Commercial shutter glass 3D systems typically run at 50 Hz to 60 Hz in order to avoid the perception of an unpleasant flicker at lower frequencies, which are below the so called *flicker fusion threshold* [De Lange Dzn, 1958; Shady et al., 2004]. Behavioural studies about visual perception found that the flicker fusion threshold varies interindividually and depends on physical properties of the stimulus, such as its brightness [Wilkins, 1995; Berman et al., 1991]. The binocular flicker fusion threshold is lower for an alternate (as it is the case for 3D shutter glasses) than for a synchronous stimulation of the two eyes [Sherrington, 1904; Baker, 1970]. In classical neurophysiological studies, the eyes were stimulated with intermittent light and brain activations were detected in the EEG that reflected the flickering light sources below the flicker fusion threshold and beyond [Adrian and Matthews, 1934; Toman, 1941; Brundrett, 1974; Lyskov et al., 1998; Herrmann, 2001] – in single subjects up to 15 Hz above the flicker fusion threshold [Porbadnig et al., 2011]. In the electroretinogram (ERG), effects of flickering stimuli were found up to 162 Hz [Berman et al., 1991], i.e. up to frequencies much higher compared to the EEG. Apparently, the visual system consists of a series of temporal low-pass filters [Shady et al., 2004] and, therefore, it is crucial to separate EEG from ERG signals.

Adding to these investigations, specifically the effect of 3D shutter glasses on neural processes during image viewing was quantified in the study presented here. It was investigated to which extent the visual cortex reflects the oscillations of the shutter glasses, in particular at high frequencies, where a flicker is no longer consciously noted. This additional processing effort is unnecessary and possibly causes visual fatigue on the long-term [Ukai and Howarth, 2008; Wilkins, 1995]. It was sought to determine the critical shutter frequency at which the ‘neural flicker’ vanishes, in order to give a recommendation to increase the shutter frequency accordingly. In this way, neural workload and the risk of visual discomfort can potentially be reduced and the usability be improved.

The following method to detect effects of 3D shutter glasses on the visual cortex is proposed. Independent component analysis (ICA) and the selection of components according to their spectral and topographic properties allows to focus on the visual cortex and to filter out undesired artefacts from non-cortical sources, such as signals originating in the retina. Quan-

tification of the impact of the shutter glasses is made possible by decoding the state of the shutter from spatio-temporal EEG features. By employing this multivariate machine learning technique on single-trial basis, the full spatio-temporal signature is exploited and allows to better detect and quantify even subtle effects in the neural data than standard univariate measures such as amplitude differences in class-wise averaged event-related potentials. This applies also to mass-univariate approaches, which moreover have to be corrected for multiple testing.

The remainder of the chapter is structured as follows. First, the experimental study is described in the sections 3.2.1 to 3.2.5 and the proposed analysis method is detailed in section 3.2.6. Then, the results of the study are presented in section 3.3, and, finally, discussed in section 3.4.

The chapter is based on the following publication:

Wenzel, M. A., Schultze-Kraft, R., Meinecke, F. C., Cardinaux, F., Kemp, T., Müller, K.-R., Curio, G., and Blankertz, B. (2016d). EEG-based usability assessment of 3D shutter glasses. *Journal of Neural Engineering*, 13(1):016003. doi: 10.1088/1741-2560/13/1/016003
© IOP Publishing. Reproduced with permission.

3.2 Material and methods

3.2.1 Experimental design

Twenty-three participants repeatedly viewed a single 2D image wearing 3D shutter glasses, while EEG was recorded. The shutter frequency was varied at ten different frequencies that had been selected based on the individual flicker fusion threshold (see sections 3.2.2 and 3.2.3). The frequencies were presented twenty times each, in random order. After every stimulus presentation for 10 s, the participants reported whether they had perceived a flicker or not (see figure 3.1 for more details).

A 2D image was presented in order to avoid potential flaws in the 3D image itself as confounding factor. In 2D images, the same image is presented to the left and to the right eye. In 3D images, left and right eye image differ and are displayed alternately on the screen. Crosstalk is a leakage between the stereo channels of 3D images [cf. Woods, 2010], leads to uncomfortable “ghosting”, increases with the display refresh rate and hence with the shutter frequency and thus would interfere with the investigation. By presenting a 2D image, the investigation was narrowed down to the effect of the oscillations of the shutter glasses on the brain, while leaving aside effects related to the display.

3.2.2 Selection of the shutter frequencies

Before the main experiment, ten shutter frequencies were selected for each participant, that captured both the critical area around the individual flicker fusion threshold (see section 3.2.3)

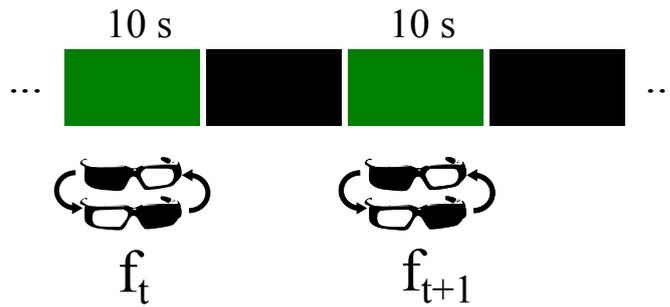


Figure 3.1: Experimental design. In a single stimulus presentation, the image was presented for 10 s, while the shutter glasses oscillated at f_t . Then, the screen turned black and the participants were requested to report with a key press if they had perceived a flicker or not. Subsequently, the shutter frequency was changed, while the screen remained black for 2 s more. Then, the next stimulus presentation with the new frequency f_{t+1} started and the image reappeared. Note that the 2D image itself did not oscillate as it would be the case for a 3D image. The image depicted a jungle (here represented by a green rectangle) and was chosen because of the homogeneous distribution of spatial frequencies. [Figure from Wenzel et al., 2016d, © IOP Publishing. Reproduced with permission. All rights reserved.]

as well as frequencies well below and above the threshold:

- Flicker fusion threshold: f_4
- Subthreshold frequencies: $f_{1,2,3} = 39 \text{ Hz} + i \times (f_4 - 39 \text{ Hz}); i \in 0, 0.4, 0.7$
- Suprathreshold frequencies: $f_{5,6,\dots,10} = f_4 + j \times (97 \text{ Hz} - f_4); j \in 0.1, 0.2, 0.4, 0.6, 0.8, 1$

All frequencies were rounded to odd numbers for three reasons: (1) to avoid the 50 Hz power line interference, (2) to avoid the overlap of a low frequency's harmonic and a high frequency, and (3) to limit the number of manual frequency settings that is restricted by the graphics card.

3.2.3 Determination of the individual flicker fusion threshold

Initially, the individual flicker fusion threshold was determined with the staircase method [Levitt, 1971]. The stimulus was presented for 4 s at a test frequency f and the participants reported whether they had perceived a flicker or not. In case of perception, the shutter frequency was increased by 2 Hz, otherwise decreased by 2 Hz. Due to this rule, the frequency approaches the flicker fusion threshold and traverses it repeatedly, while remaining close to it. The procedure was repeated until the perception report had switched from 'yes' to 'no' (or vice versa) twenty times. The initial shutter frequency was set to be 49 Hz (from prior knowledge this is close to the flicker fusion threshold). To give a frame of reference, catch stimuli with either an easily detectable flicker at a frequency of 41 Hz, or a non-detectable flicker at a frequency of 85 Hz

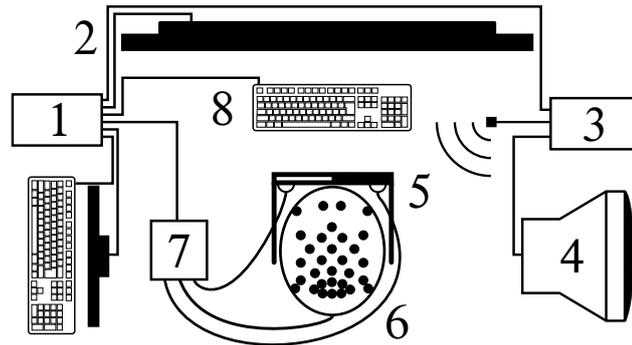


Figure 3.2: Experimental setup. **1** Computer for experimental control and data acquisition. **2** LCD television. **3** Computer with infrared light emitter for shutter glasses control. **4** Cathode ray tube monitor. **5** Shutter glasses with attached photodiodes. **6** EEG cap. **7** Amplifier and analogue digital converter of EEG and photodiode signals. **8** Keyboard to enter if a flicker was perceived or not. [Figure from Wenzel et al., 2016d, © IOP Publishing. Reproduced with permission. All rights reserved.]

were presented with a chance of one in eight each. The flicker fusion threshold was obtained by averaging all frequencies presented, discarding catch stimuli and stimuli before the first ‘yes’/‘no’ transition of the flicker report.

3.2.4 Data acquisition

Experiments with twenty-three participants with normal or corrected to normal vision and no report of eye or neurological diseases were conducted in accordance with the local statutory requirements. The subjects gave their informed written consent to participate in the experiment. The age of the nine women and fourteen men ranged from 19 to 54 years and was on average 29.2 ± 7.1 years (mean \pm standard deviation). The participants gave their informed written consent to take part in the study. EEG was recorded with 32 active electrodes (‘actiCAP’, ‘BrainAmp’, ‘BrainVision Recorder’; Brain Products, Germany), the ground electrode was located on the forehead and two electrodes on left and right mastoids served as references. The electrode montage followed the international 10–20 system and had a high electrode density over the visual cortex: Fp1, Fp2, F9, Fz, F10, FC5, FC1, FC2, FC6, C3, Cz, C4, CP5, CP1, CP2, CP6, P9, P7, P3, Pz, P4, P8, P10, PO7, PO3, POz, PO4, PO8, O1, Oz, O2. Photodiodes were mounted behind the shutter glasses and luminance time-series were recorded to track the timing of the opening and closing of the shutter.

3.2.5 Experimental setup

The experimental setup allowed for remotely adjusting the various shutter frequencies (see figure 3.2): A computer displayed the image on a LCD TV monitor (101.7 cm x 57.2 cm, 1920 px

Chapter 3. EEG-based usability assessment of stereoscopic displays

x 1080 px; *Bravia KDL-46HX805*; Sony, Japan) and recorded the EEG and the flicker report entered on the keyboard. A second computer controlled with an infrared emitter the 3D shutter glasses (*3D Vision Kit 2*; NVIDIA, USA) and changed the shutter glasses' frequency, after receiving a command from the first computer.

The functioning of the shutter glasses was enabled by the program *Stereoscopic Player* (3dtv.at, Austria) and the shutter glasses' frequency was adjusted using the program *QRes* (Engelbrecht, the Netherlands). Because the image was displayed in 2D and not in 3D, a synchronization of shutter glasses and TV was not necessary. For mere technical reasons, a cathode ray tube monitor (Iiyama, Japan) was connected to the second computer (because it allowed to adjust arbitrary and high screen refresh rates and, thus, shutter frequencies).

Beforehand, the proper functioning of the shutter glasses was checked with an oscilloscope. The close viewing distance of 1 m ensured that the TV screen covered a large part of the visual field, such that the especially motion-sensitive peripheral areas of the retina were stimulated, too [Hartmann et al., 1979]. The visual angle subtended by the screen was 54° in horizontal and 32° in vertical direction. During the experiment, the windows of the room were darkened, lights were switched off, and the screen of the recording computer was covered with a piece of fabric in order to avoid light sources other than the display screen.

3.2.6 Data analysis

Neural correlates of the flicker were extracted from the EEG using ICA, and the impact of the flicker on the brain was quantified for each frequency by decoding the state of the shutter glasses from the neurophysiological data. The individual steps of the analysis procedure are explained in the following:

Preprocessing. Raw data sampled at 1000 Hz were re-referenced to the linked mastoids. Slow drifts were removed by subtracting a moving average over 1500 ms. The EEG data recorded during epochs of stimuli presentation were assembled.

ICA to extract neural correlates of the flicker. Each cortical source is not measured in a single EEG electrode, but exhibits a characteristic field pattern, which superimposes linearly and leads to the complex dynamics of the measured EEG. In addition, the ERG and undesired artefacts interfere with the signals that originate in the cortex. ICA is a statistical technique for extracting both the source signals and the corresponding field patterns from the recorded EEG data and thereby allows for separating visually evoked cortical signals that are of interest here from all other cortical and non-cortical signals (cf. section 2.3.2).

The source separation algorithm TDSEP [Ziehe and Müller, 1998; Ziehe et al., 2000] was applied along the temporal dimension of the multichannel EEG data of each participant, including epochs of all shutter frequencies. Spectra were computed for each component, separately for periods of each shutter frequency. One to five components were selected manually whose spectra revealed prominent peaks corresponding to the shutter frequencies or their harmonics.

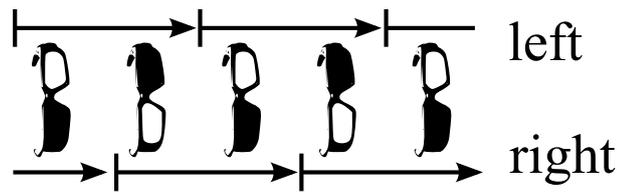


Figure 3.3: Epochs of the class ‘left’ aligned to the opening of the left shutter glass, epochs of the class ‘right’ aligned to the opening of the right shutter glass. The length of each epoch corresponded to the respective wave length of the shutter. [Figure from Wenzel et al., 2016d, © IOP Publishing. Reproduced with permission. All rights reserved.]

Criterion for exclusion was if the field pattern had large weights at frontal electrode locations such that the source could not be attributed to occipital or parietal brain areas where the visual cortex is situated.

Quantification of the ‘neural flicker’ with classification. The impact of the flicker on the brain was quantified for each shutter frequency f by classifying short epochs of the selected independent component(s). The epochs were as long as the shutter wavelength T and were aligned either to the shutter of the left or the right glass (see figure 3.3). In the next section, it is explained how the epochs were extracted. Classifiers were trained with linear discriminant analysis (LDA; cf. section 2.3.3.3) in order to discriminate the epochs of the two classes. The classification performance was evaluated for every shutter frequency with a 10-fold cross-validation (cf. section 2.3.3.6).

Notably, ICA was not used to increase classification performance, but to ensure that only information from cortical sources is used, see the discussion in section 3.4.4 below.

Extraction of the samples of the two classes. The samples used for the classification detailed in the previous section were extracted as follows: The selected independent component(s) were split into time intervals recorded during single stimulus presentations. Markers were set every half the shutter wavelength $T/2 = 1/(2 \times f)$, such that they were alternately aligned to the left and the right shutter. Each shutter frequency was presented twenty times (compare section 3.2.1). However, the phase of the shutter was different in these twenty repetitions. Hence, the phase was set to the same value beforehand: for each stimulus presentation, the delay between the signal of the left photodiode (see section 3.2.4) and a sine wave of the respective frequency was determined by cross-correlation and the markers were shifted according to the delay. After every marker, an epoch as long as the shutter wavelength T was cut out (see figure 3.3) and all epochs from the twenty stimulus presentations of one frequency were assembled.

Assessing the classification performance. The area under the curve (AUC) of the receiver operating characteristic (ROC) was computed to assess the performance of the binary classification [Fawcett, 2006]. The AUC indicates the discriminability of the short epochs after opening the left versus opening the right glass and reflects the effect of the shutter glasses on

Table 3.1: The ten selected shutter frequencies averaged over all participants [Hz]. f_4 corresponds to the flicker fusion threshold. The small difference between f_4 and the threshold results from the rounding to odd numbers as motivated in section 3.2.2.

f_1	f_2	f_3	f_4	f_5
39.0 ± 0	42.2 ± 1.2	44.7 ± 2.3	47.7 ± 3.2	52.5 ± 2.5
f_6	f_7	f_8	f_9	f_{10}
57.2 ± 2.5	67.2 ± 1.8	77.3 ± 1.5	87.1 ± 0.4	97.0 ± 0

the neural processes. The more the AUC differs from the chance level of 0.5, the larger is the impact of the oscillations of the shutter glasses on the brain.

Classification of the original EEG data without ICA. In an additional analysis, the ICA step was skipped and the original EEG data were used for the classification to identify the effect of the neural source extraction. All other analysis steps remained unchanged.

Event-related potentials ‘left’ versus ‘right’. In order to demonstrate that the extraction of the source of the ‘neural flicker’ with ICA is essential for this analysis, bi-serial correlation coefficients (cf. section 2.3.3.7) of EEG epochs and class labels (left=1, right=-1) were computed for 39 Hz and for 97 Hz and averaged over the first and the second half of the epochs, i.e. $[0, T/2]$ and $]T/2, T]$.

3.3 Results

3.3.1 Flicker fusion thresholds and shutter frequencies

The flicker fusion thresholds, determined before the main experiment, ranged from 42.6 Hz to 52.3 Hz with an over all average of 47.4 ± 3.0 Hz (mean \pm standard deviation). The ten selected shutter frequencies are presented in table 3.1.

3.3.2 ICA to extract neural correlates of the flicker

ICA revealed for twenty of the twenty-three participants at least one independent component with prominent peaks in the power spectrum at frequencies corresponding to the shutter frequencies used in the experiment. The peaks were found in particular at lower frequencies. Components with this spectral property, typically had scalp patterns that showed the most distinctive values over occipital and parietal scalp locations (see figure 3.4).

3.3.3 Quantification of the ‘neural flicker’ with classification

The results of the classification of selected independent components using LDA are depicted in figure 3.5 (single participants) and in figure 3.6 (averages across participants). It can be

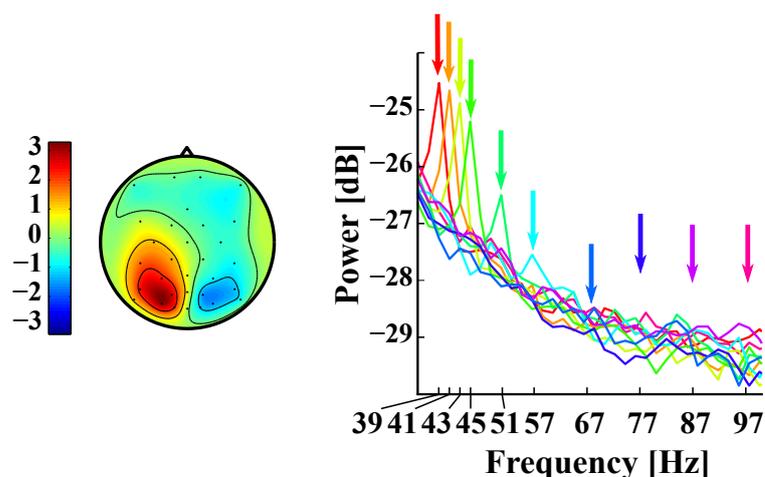


Figure 3.4: The independent component selected as neural flicker correlate based on its spectral and topographic properties (representative participant *VPjau*). *Left*: Colour-coded scalp topography (nose up) of the component's mixing weights. Large absolute values are prominent at electrodes close to the visual cortex. *Right*: The component's frequency spectrum was estimated separately for the ten shutter frequencies. Each arrow indicates with its horizontal position the shutter frequency relevant for the spectrum of its colour. Spectral peaks corresponded to the respective shutter frequency up to 67 Hz. The gaps between the arrows at 77 Hz, 87 Hz, and 97 Hz and the spectra illustrate the absence of corresponding spectral peaks. [Figure from Wenzel et al., 2016d, © IOP Publishing. Reproduced with permission. All rights reserved.]

observed that the AUC decreases with increasing shutter frequency. On average, the AUC was significantly better than the chance level for frequencies up to $f_7 = 67.2 \pm 1.8$ Hz ($Z = 3.83$, $p < 0.05$, one-tailed Wilcoxon signed-rank test).

As a control analysis, a permutation test was performed which showed that the AUC was not significantly better than chance when the labels were randomly shuffled ($p > 0.05$).

While the lengths of the epochs decreased with the shutter frequency, the number of available samples increased. Hence, an additional classification procedure was performed using an equal set size and epoch length for all frequencies. Although the AUC showed a small decrease in comparison to the procedure with a variable set size and epoch length, it remained significantly above chance up to frequency level f_7 ($Z = 3.71$, $p < 0.05$).

3.3.4 Behavioural data

The proportion of experimental stimuli for which the participants reported to have perceived a flicker is represented by the grey line in figure 3.6. For frequencies higher than the flicker fusion threshold (f_4), the flicker detection rate dropped rapidly and was merely 9.6 % at f_6 .

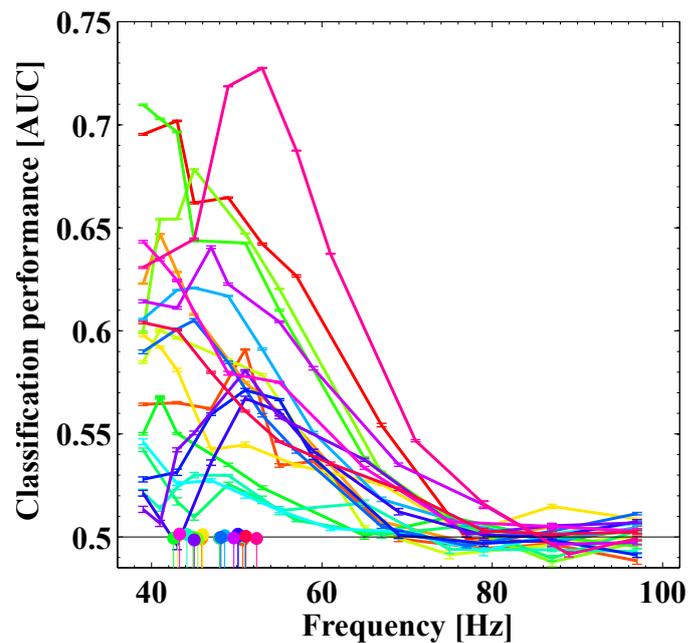


Figure 3.5: Classification of short epochs of selected independent components. The AUC-scores of the twenty participants are presented for the different shutter frequencies. Error bars represent the standard deviation. Coloured discs indicate the individual flicker fusion thresholds determined before the experiment. The individual classification performances of neighbouring shutter frequencies were significantly correlated up to the pair of f_7 and f_8 ($r = 0.49$, $p < 0.05$). [Figure from Wenzel et al., 2016d, © IOP Publishing. Reproduced with permission. All rights reserved.]

3.3.5 Classification of the original EEG data without ICA

In a parallel analysis, the ICA step was skipped and short epochs of the *original EEG data* were classified. On average, classification better than chance was possible at all ten frequency levels including $f_{10} = 97$ Hz (cf. figure 3.7 and also section 3.3.6 with figure 3.8).

3.3.6 Event-related potentials ‘left’ versus ‘right’

The difference between the two classes (‘left’ and ‘right’) was assessed with bi-serial correlation coefficients of EEG epochs and class labels. The resulting values are displayed as scalp topographies in figure 3.8. They illustrate that the two classes differed for shutter frequencies of 39 Hz at occipital and parietal electrodes close to the visual cortex, but for 97 Hz at frontal electrodes close to the eyes and the shutter glasses.

3.3.7 Critical frequencies 67 Hz and 77 Hz

Effects of the shutter glasses on the brain were detected up to $f_7 = 67.2 \pm 1.8$ Hz in contrast to higher shutter frequencies starting with $f_8 = 77.3 \pm 1.5$ Hz (cf. section 3.3.3, figure 3.5 and

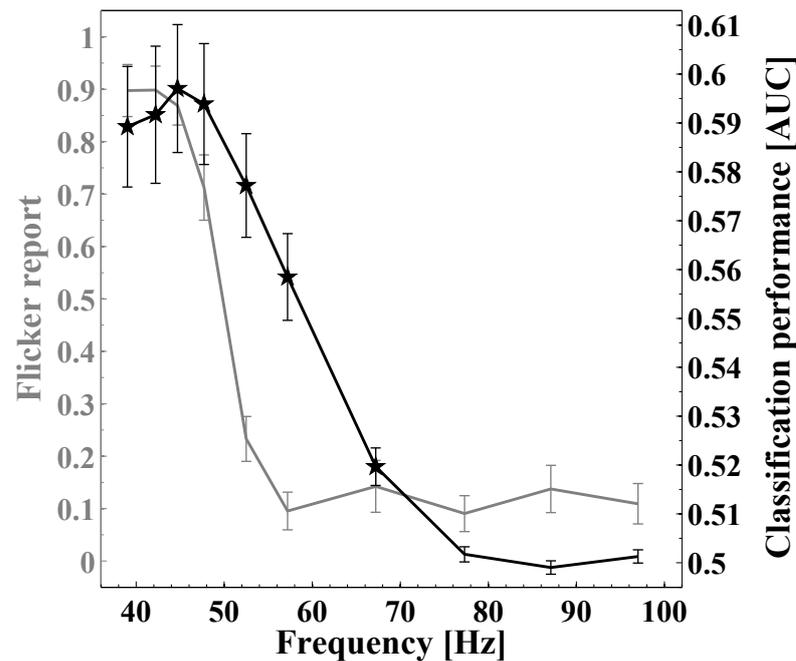


Figure 3.6: Averages across participants. *Black*: Classification of short epochs of selected independent components. AUC-scores were averaged across participants separately for each shutter frequency level. Stars indicate AUC-scores significantly above chance level ($p < 0.05$, one-tailed Wilcoxon signed-rank test). Classification results were significant up to frequency level $f_7 = 67.2$ Hz, and thus around 20 Hz above the flicker fusion threshold of 47.4 Hz. *Grey*: Average flicker detection rates during the experiment, i.e. the proportions of experimental stimuli for which the participants reported a flicker.

Error bars represent the standard error of the mean. Although the exact frequency values depend on the participant-specific flicker fusion thresholds, error bars along the x-axis were omitted due to the small variations (see table 3.1). [Figure from Wenzel et al., 2016d, © IOP Publishing. Reproduced with permission. All rights reserved.]

figure 3.6). To inspect the individual sensitivities at these apparently critical frequencies in more detail, the results of the single subjects for f_7 were compared with those for f_8 in figure 3.9.

3.4 Discussion

3.4.1 Effects of the shutter glasses on the visual cortex and the human perception

Twenty of the twenty-three participants featured (an) independent component(s) with characteristic spectral and topographic properties that can be considered as neural correlates of the flicker, i.e. neural activity that reflects the oscillations of the shutter glasses and that probably originates in the visual cortex (compare figure 3.4). By decoding the state of the shutter glasses from these components, the impact of the shutter on the visual cortex was quantified.

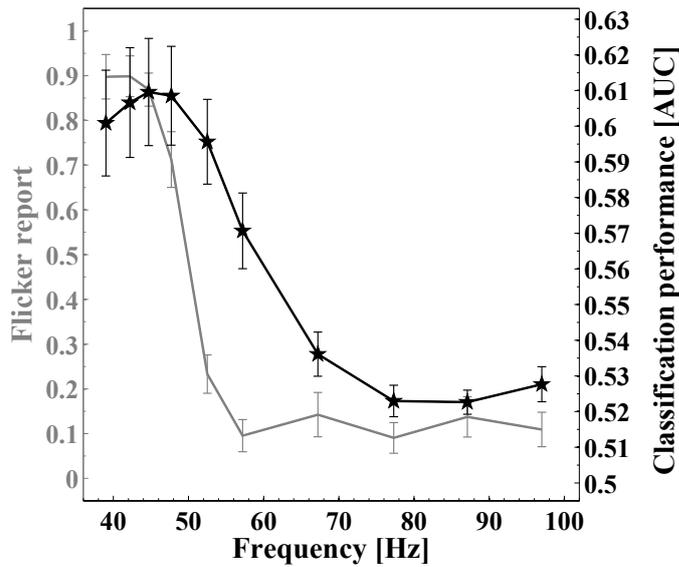


Figure 3.7: Classification results for the original EEG data without ICA. [Figure from Wenzel et al., 2016d, © IOP Publishing. Reproduced with permission. All rights reserved.]

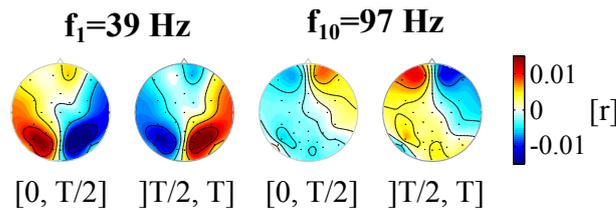


Figure 3.8: Bi-serial correlation coefficients of EEG epochs and class labels (*left*=1, *right*=-1) were computed for 39 Hz and for 97 Hz and averaged over the first and the second half of the epochs, i.e. $[0, T/2]$ and $]T/2, T]$ (participant *VPjav*). [Figure from Wenzel et al., 2016d, © IOP Publishing. Reproduced with permission. All rights reserved.]

Neural correlates of the flicker were traced up to a shutter frequency of 67.2 Hz, which is about 20 Hz over the perception threshold of the participants. For higher frequencies – the next tested frequency was around 77.3 Hz – no ‘neural flicker’ was detected in the data. The results suggest that at this frequency the neural processing of the flicker decreases significantly and that an unnecessary workload can be avoided by setting the shutter to frequencies of 77.3 Hz and above. The risk of visual discomfort could thereby be reduced and the usability of 3D shutter glasses improved. The analysis of neurophysiological data proved to be more sensitive than directly asking the participants, which shows the additional benefit of employing BCI-based usability assessment tools on top of test person reports. This observation was made before with respect to other application areas [e.g. Porbadnigk et al., 2011, and Kohlmorgen et al., 2007]. Note that the class labels were derived from optical measurements of the left and the right shutter, and not from behavioural responses of the participants, which could result in a ‘systematic label noise’ [Porbadnigk et al., 2015].

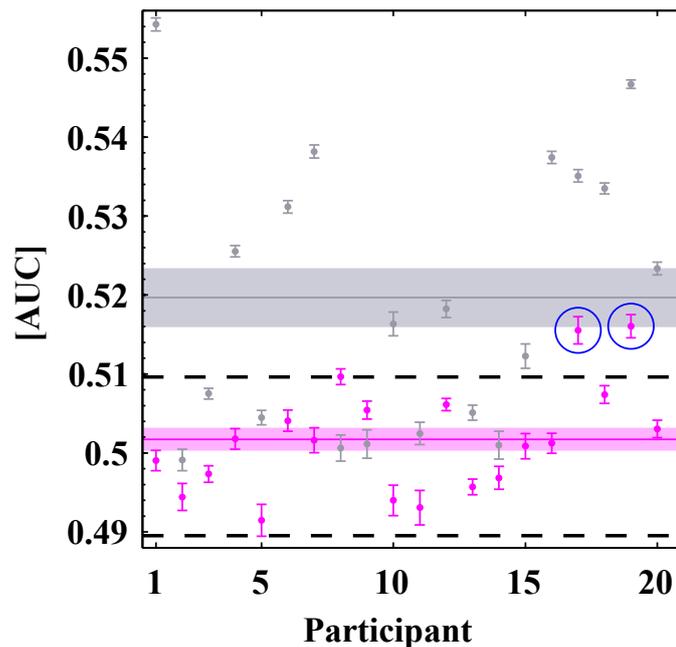


Figure 3.9: Individual classification results for the critical frequencies f_7 (grey) and f_8 (magenta). Lines and shades illustrate the respective mean and standard deviation for all participants. The dashed black lines indicate the 95 % confidence interval of a random classifier estimated from the results of classifications with shuffled labels (cf. section 3.3.3). The results of two participants for f_8 (blue circles) were comparably high, they surpassed the upper confidence limit of the random classifier and were close to the average result for the lower frequency f_7 . [Figure from Wenzel et al., 2016d, © IOP Publishing. Reproduced with permission. All rights reserved.]

The critical frequencies $f_7 = 67.2$ Hz and $f_8 = 77.3$ Hz were inspected in more detail (cf. section 3.3.7 and figure 3.9). Two participants featured comparably high classification results for f_8 that were close to the average result for the lower frequency f_7 . This finding can either be explained by mere random variations of the individual values around the population mean or because there are indeed individual variations in the flicker sensitivity. The latter case is supported by a significant correlation of the individual classification performances for neighbouring shutter frequencies up to the pair of f_7 and f_8 ($r = 0.49$, $p < 0.05$; cf. figure 3.5). Apparently, some persons are particularly sensitive and require higher shutter frequencies than the majority to avoid the ‘neural flicker’. The found inter-subject variability matches the reports in the literature [e.g., Brundrett, 1974; Lyskov et al., 1998; Porbadnigk et al., 2011].

3.4.2 Behavioural flicker fusion threshold

When the frequency of an oscillating light source crosses the flicker fusion threshold, the perception passes from ‘flickering’ to ‘continuous’ or vice versa. The threshold was determined to be 47.4 ± 3.0 Hz with the staircase method that estimates the stimulus level where the

detection rate is 0.5 [Levitt, 1971]. However, the average rate of behavioural report at f_4 was with 0.71 higher than the expected value of 0.5, which shows that the participants were more sensitive towards the detection of the flicker during the main experiment, as compared to the prior determination of the flicker fusion threshold. Nevertheless, the difference is negligible, because a detection rate of 0.5 was reached approximately at 49.8 Hz (interpolation between f_4 and f_5) and at $f_{5,\dots,10}$ the detection rates were significantly below 0.5 ($Z = -3.51$, $p < 0.05$, one-tailed Wilcoxon signed-rank test).

3.4.3 Why ICA is essential for this analysis

The selection of the source of the ‘neural flicker’ with ICA is a crucial step in the proposed analysis procedure, even though skipping the ICA step leads to significant classification results at all ten frequency levels including $f_{8,9,10}$ (cf. section 3.3.5 and figure 3.7). The correlation of EEG epochs and class labels (see figure 3.8) clearly demonstrates that classifying the original EEG data would not reflect cortical processes and cannot be regarded as a measure for the ‘neural flicker’. At high shutter frequencies, signals at frontal electrodes that probably originate in the retina or in the shutter electronics are informative about the classes. Electroretinographic investigations have shown that the human retina can reproduce a rhythm of up to 162 Hz [Berman et al., 1991]. Using ICA, it was possible to specifically extract those EEG components that could be attributed to neural activity in the visual cortex, and discard signals originating at other locations (cf. figure 3.4).

3.4.4 Assumption and limitation

The assumption was made that the neural sources that process the flicker remain the same for all shutter frequencies. Thus, ICA was applied to the whole data set of a participant containing all ten shutter frequencies and before the cross-validation procedures – even though the latter violates the complete separation of test and training set [cf. Lemm et al., 2011]. Again, it should be pointed out that ICA was used to restrict the analysis to cortical sources and not to improve the classification performance. Quite the contrary, classifiability diminished when ICA was used. A limitation of the presented approach is that the classification is only based on lateralised effects (left vs. right eye signals).

3.4.5 Relation between neurally-detectable flicker and long-term user satisfaction

A basic question that concerns all investigations, where neurophysiological techniques achieved results more sensitive than behavioural methods, is: does the detectable cortical effect have a relevant impact on the problem addressed? Therefore, it is important to investigate the effect of an increased shutter glasses frequency on the prevalence of visual discomfort and on the long-term user satisfaction in future studies. While a shutter frequency as high as possible is

desirable to avoid unnecessary neural workload, possible side-effects of an increased shutter frequency, e.g., stronger crosstalk or a higher price have to be taken into account to find the frequency with the optimal trade-off.

3.4.6 Conclusion

The problem was addressed of the neural processing load imposed on the viewer of stereoscopic television by the shutter glasses. It was demonstrated that methods from BCI can contribute to the usability assessment by providing objective measures of the neural effort. Usability impediments were revealed that were not detected by the test subjects because of the limits of human perception. Finally, the assessment made it possible to recommend changes to optimise the usability and, ideally, the long-term user satisfaction.

3.5 Lessons learned

- BCI methods provided an objective and particularly sensitive measure of the workload that a device demands from the brain of the user. Usability flaws were discovered that test persons could not notice due to the limits of human perception.
- The neural processing effort imposed by shutter glasses on the viewer of stereoscopic television was quantified as a function of the shutter frequency. At low shutter frequencies, an annoying and fatiguing flicker is perceived, which vanishes above a critical frequency. For optimal viewing comfort, shutter glasses run at a frequency above the perception threshold.
- The impact of the shutter glasses on the brain was quantified with multivariate data analysis techniques applied to the EEG. Independent component analysis made it possible to focus on the visual cortex. The effect on the brain was quantified by decoding the state of the shutter from the EEG with linear discriminant analysis.
- Effects of the shutter glasses on the brain were detected also for common shutter frequencies, and up to about 20 Hz above the flicker perception threshold.
- Increasing the shutter frequency can potentially avoid an imperceptible neural strain and prevent visual fatigue.

4 Real-time inference of word relevance from EEG and eye gaze

4.1 Introduction

Electromagnetic fields of the brain and eye movements may carry information about the subjective relevance of the single items present in the visual surrounding. This implicit information can potentially be decoded in real-time in order to infer the current interest of the individual person. Previous research on brain-computer interfacing (BCI) has shown that it can be estimated which stimuli aroused the interest, when a stimulus sequence is viewed – by detecting multivariate patterns in non-invasive recordings of the brain activity.

However, familiar stimuli are typically presented again and again in BCI, and can therefore be easily recognised, regardless of whether they are letters, pictures of faces, geometric shapes or merely colours [e.g. Farwell and Donchin, 1988; Kaufmann et al., 2011; Treder et al., 2011; Acqualagna et al., 2013; Acqualagna and Blankertz, 2013; Seoane et al., 2015]. In contrast, the regular visual environment contains items that have to be interpreted with respect to their meaning, most notably words in the case of written text. The interpretation of the semantics goes beyond the simple recognition of a previously known letter, picture, or shape that is repeatedly flashed.

Accordingly, the question was addressed if the relevance inference from the electroencephalogram (EEG) can be also applied in settings where semantic content has to be interpreted. Readers looked for words belonging to one out of five semantic categories, while a stream of words passed at different locations on the screen (cf. figure 4.1). The words were dynamically replaced (when they had been fixated with the eye gaze) by new words fading in. It was estimated in real-time during the experiment which words and thus which semantic category interested the reader, based on information implicitly contained in the EEG and eye tracking signals. The estimates were visualised for demonstration purposes on the edge of the screen, and were updated as soon as a new word had been read. In this way, the reader could learn about the current estimates (for each of the five categories), and could observe how evidence was accumulated over time. Prior to the online inference (cf. section 4.2.2), a classifier had to be trained to estimate the word relevance based on the signals (cf. section 4.2.1).

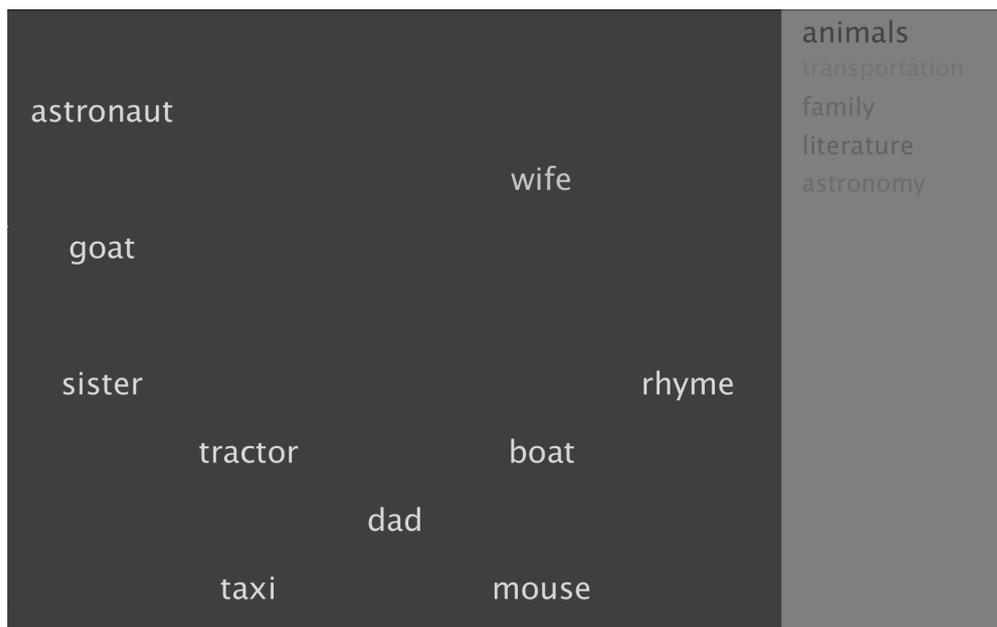


Figure 4.1: Looking for words (left) related to one out of five semantic categories (right). The current estimates are represented by the luminances of the five category names. [Figure from Wenzel et al., 2017, © IOP Publishing. Reproduced with permission. All rights reserved.]

In contrast to recent investigations with similar objectives [Geuze et al., 2013, 2014; Eugster et al., 2014, 2016], several words were displayed at the same time on the screen. The participants could scan the words without restrictions on the eye movements. Neural activity was related with eye tracking to the respective word looked at, like in studies on reading [e.g. Baccino and Manunta, 2005; Dimigen et al., 2011, 2012; Kliegl et al., 2012; Kornrumpf et al., 2016] and on visual search, which have shown that sought-for items evoke a detectable neural response when they are fixated with the eye gaze [cf. Kamienkowski et al., 2012; Brouwer et al., 2013; Kaunitz et al., 2014; Kauppi et al., 2015; Golenia et al., 2015; Wenzel et al., 2016c; Ušćumlić and Blankertz, 2016; Finke et al., 2016].

The subjective relevance of the visual surrounding can be mapped with this approach by assigning relevance scores to the single items in view. The obtained information can potentially be exploited for the optimisation of websites or stores, or for the usability evaluation of aircraft or car cockpits. Besides, BCI-based relevance maps make it possible to conduct new kinds of experiments in basic research. Furthermore, the obtained information can be aggregated in order to estimate the current interest of the individual person. The resulting dynamic user interest profile would render possible novel types of adaptive software and personalised services, that enrich the interaction between human and computer by adding implicit information to the explicit interaction [cf. Eugster et al., 2014, 2016; Kauppi et al., 2015; Finke et al., 2016; Pohlmeier et al., 2011; Zander and Kothe, 2011; Ušćumlić et al., 2013; Jangraw et al., 2014; Blankertz et al., 2016]. Less obtrusive and more convenient EEG systems with sufficient signal quality are prerequisite for the application in practice [cf. Nikulin et al.,

2010; Looney et al., 2014; Debener et al., 2015; Norton et al., 2015; Goverdovsky et al., 2016a,b].

The chapter is based on the following publication.

Wenzel, M. A., Bogojeski, M., and Blankertz, B. (2017). Real-time inference of word relevance from electroencephalogram and eye gaze. *Journal of Neural Engineering*. doi: 10.1088/1741-2552/aa7590

© IOP Publishing. Reproduced with permission.

4.2 Materials and methods

4.2.1 Calibration

Labelled EEG and eye tracking data were recorded in order to train a classifier that could predict the relevance of the single words in the subsequent online phase (cf. section 4.2.2). The participants selected one out of five given semantic categories. Subsequently, twenty-two words were drawn randomly from the five categories, with a contribution of 20 % per category on average. Words faded in on the screen at predefined positions in random order (cf. figure 4.1), and were faded out when they had been fixated with the eye gaze (with a delay of one second).

Examples of the categories and words are:

- Astronomy: orbit, galaxy, universe, meteorite.
- Time: future, seconds, hourglass, minute.
- Furniture: bathtub, closet, stool, bed.
- Transportation: taxi, canoe, tractor, helicopter.
- Visual art: palette, pencil, sculpture, crayon.

The participants were requested to remember the words that belonged to the chosen category. When the participants had looked at all words, they were asked to recall the relevant words from their memory. For this purpose, the words reappeared truncated (to about 40 % of the original number of letters) at shuffled positions. Relevant words had to be selected with the mouse. Subsequently, the accuracy of the recall was checked and reported. This procedure helped to involve the participants in the task, but avoided interference of motor activity during the acquisition of the EEG data.

For the study, a corpus had been generated of seventeen semantic categories with twenty words each, both in English and German depending on the language skills of the participant (cf. section 4.3.1). The seventeen categories were: animals, furniture, transportation, body parts, family, food, literature, country names, astronomy, music, finance, buildings and structures,

healthcare, sports, time, clothes, and visual art. The calibration phase consisted in seventeen blocks with four repetitions each. At the beginning of each block, a semantic category (out of five options) could be chosen. The categories offered for selection changed during the course of the experiment, such that each of the seventeen categories could serve once as category of interest. During the recording, it was tracked which category had been chosen by the participant and thus which single words were relevant.

Feature vectors were extracted from the recorded EEG and eye tracking data with the intention to capture processes related to word reading and categorisation (details below). The feature vectors were labelled depending on whether the word fixated at this moment was relevant or irrelevant to the chosen category of interest. Subsequently, a classification function was trained with regularized linear discriminant analysis [Friedman, 1989] to discriminate the feature vectors of the 'relevant' and the 'irrelevant' class [Blankertz et al., 2011]. The shrinkage parameter was calculated with an analytic method [Ledoit and Wolf, 2004; Schäfer and Strimmer, 2005].

4.2.1.1 Feature extraction

The multi-channel EEG signal was re-referenced to the linked mastoids and low-pass filtered (with a second order Chebyshev filter; 42 Hz pass-band, 49 Hz stop-band). The continuous signal was segmented by extracting the interval from 100 ms to 800 ms after the onset of every eye fixation. Slow fluctuations in the signal were removed by baseline correction (i.e. by subtracting the mean of the signal within the first 50 ms after fixation onset from each epoch). The signal was downsampled from the original 1000 Hz to 20 Hz in order to decrease the dimensionality of the feature vectors to be obtained (14 values per channel). A low dimensionality in comparison to the number of available samples is beneficial for the classification performance, because the risk of overfitting to the training data is reduced [Blankertz et al., 2011]. The multi-channel signal was vectorised by concatenating the values measured at the 62 scalp EEG channels at the 14 time points resulting in a $62 * 14 = 868$ dimensional vector per epoch. The fixation duration was concatenated as additional feature to the EEG feature vector.

Note that other eye tracking features, e.g. the gaze velocity, could not be exploited, because they are not provided in real-time by the application programming interface of the device, and that two additional EEG electrodes, which were not situated on the scalp and served for re-referencing and electrooculography, were excluded from the set of 64 electrodes in total.

4.2.2 Online prediction

The subjective relevance of words to a semantic category was inferred online with the previously trained classifier. Again, the participants read words and were asked to look for words related to one out of five semantic categories. The words faded in and out similar to the calibration phase but vacant positions were replaced by new words fading in. In this way, all

hundred words of the five involved categories were shown. Usually, several words were present on the screen at the same time. The classifier predicted online for each fixated word if it was relevant to the category of interest or not, based on the incoming EEG and eye tracking data.

The class membership probability estimates for the single words were assigned to the corresponding semantic category and all estimates obtained so far were averaged per category. The resulting five-dimensional vector indicated how likely each category was of interest. The vector was normalised to unit length, determined the font size and luminance of the visualisation of the five category names on the right side of the screen (cf. figure 4.1), and was updated when a new word had been fixated with the eye gaze. It was initialised with neutral values for the initial period when only few words had been read and not every category was captured. The participants were informed about the predictive mechanism underlying the adaptive visualisation in order to foster task engagement. A recall task like in the calibration phase (cf. section 4.2.1) was not included in view of the objective to exploit only *implicit* information. The procedure was repeated seventeen times with new combinations of five categories. At the beginning of each repetition, the participants indicated the selected category of interest for later validation, and the previously collected relevance estimates were cleared.

Remark for the sake of completeness: the classifier output was dichotomised to zero or one in the actual visualisation during the experiment. In contrast, class membership probability estimates ranging between zero and one were employed for the figures presented here.

4.2.3 Experimental setup

An apparatus was developed that allowed for making inferences from combined EEG and eye tracking data in real-time and displaying this information in an adaptive graphic visualisation.

4.2.3.1 Key constituents of the system

The system comprised an EEG device, an eye tracker, two computers and a screen that the test person was looking at (cf. figure 4.2). EEG was recorded with 64 active electrodes arranged according to the international 10–20 system (*ActiCap*, *BrainAmp*, BrainProducts, Munich, Germany; sampling frequency of 1000 Hz). The ground electrode was placed on the forehead and electrodes at the linked mastoids served as references. An eye tracker, connected to a computer (PC 1), detected eye fixations in real-time (*RED 250*, *iView X*, SensoMotoric Instruments, Teltow, Germany; sampling frequency of 250 Hz). A second computer (PC 2) acquired raw signals from the EEG device (with the software *BrainVision Recorder*, BrainProducts, Munich, Germany), and obtained preprocessed eye tracking data from PC 1 over network using the *iView X API* and a custom server written in *Python 2.7* (<https://python.org>). EEG and eye gaze data were then streamed to in-house software written within the framework of the *BBCI-Toolbox* (https://github.com/bbci/bbci_public) running in *Matlab 2014b* (MathWorks, Natick, USA). The graphic visualisation was computed with custom software written in *Processing 3*

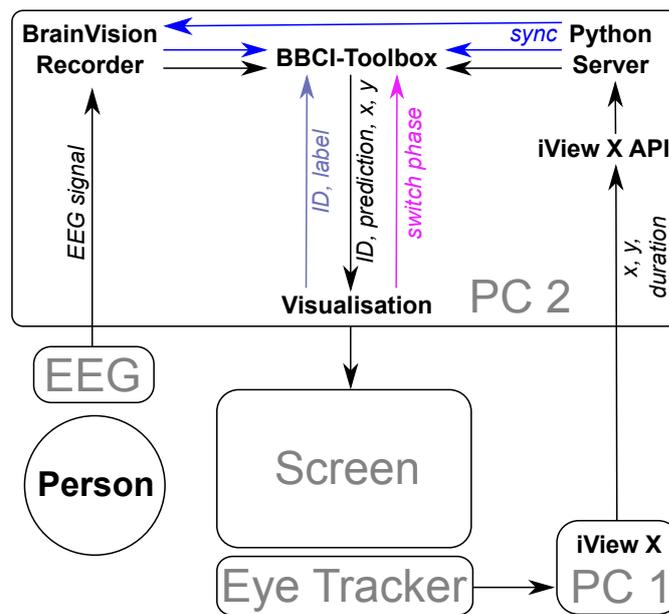


Figure 4.2: The apparatus allows for making inferences from EEG and eye tracking signals in real-time and displaying the obtained information in an adaptive graphic visualisation. [Figure from Wenzel et al., 2017, © IOP Publishing. Reproduced with permission. All rights reserved.]

(<https://processing.org>) and displayed on the screen (60 Hz, 1680 x 1050 pixel, 47.2 cm x 29.6 cm).

4.2.3.2 Synchronisation of EEG and eye tracking signals

When the data acquisition started, the Python server sent a sync-trigger into the EEG signal and transmitted the current time stamp of the eye tracker to the BBCI-Toolbox. These simultaneous markers allowed for synchronising the two measurement modalities.

4.2.3.3 Workflow of the system

The experiment included several phases, which could be switched by the visualisation software with messages sent over a TCP connection. During the calibration phase, EEG and eye tracking data were recorded to train a model (cf. section 4.2.1) that was supposed to predict the relevance of each word read by the subject in the subsequent online phase (cf. section 4.2.2). Feature vectors were extracted from the ongoing EEG and eye tracking signals, every time the eye tracker had detected a new eye fixation (cf. section 4.2.1). The visualisation software checked if the eye fixation was situated on a word displayed on the screen, according to the received x-y-coordinates. During the calibration phase, the feature vectors were labelled, depending on whether the word belonged to the category of interest or not. Labels and feature vectors were matched according to a unique identifier (ID) of each eye fixation. During the

online phase, the graphic visualisation adapted according to the incoming predictions. The architecture of the system is modular and the visualisation module can easily be replaced by other software for novel applications that depend on making real-time inferences from EEG and eye tracking signals. The communication protocol that enables the visualisation module to interact with the other parts of the system offers three types of interactions. The visualisation module can (a) switch between calibration and online phase and an initial adjustment of the eye tracker, (b) can receive relevance estimates from the BBCI-Toolbox, and (c) can mark events and stop data acquisition by sending markers into the EEG.

4.2.4 Data acquisition

Experiments with three female and twelve male participants with normal or corrected to normal vision, no report of eye or neurological diseases and ages ranging from 21 to 40 years (median of 28 years) were conducted, while EEG, eye tracking and behavioural data were recorded. Ten people performed the experiment in their mother tongue of German and five people with other first languages accomplished the task in English, which was not their mother tongue. The subjects gave their informed written consent (a) to participate in the experiment and (b) to the publication of the recorded data in anonymous form without personal information. The study was approved by the ethics committee of the Department of Psychology and Ergonomics of the Technische Universität Berlin (reference BL_03_20150109).

4.3 Results

4.3.1 Calibration

The participants recalled the words that were relevant to the category of interest with an average accuracy of 80 %, ranging from 72 % to 84 % in the individuals. Classifiers were trained individually for each participant to detect relevant words with EEG and eye tracking data recorded during the calibration phase (cf. section 4.2.1). In the subsequent online phase, the classifiers were applied to the data incoming in real-time (cf. section 4.2.2).

Additionally, the performance of the classifiers was assessed in ten-fold cross-validations using only the data recorded during the calibration phase. The area under the curve (AUC) of the receiver operating characteristic served as performance metric [Fawcett, 2006]. An AUC of 0.63 ± 0.01 (mean \pm standard error of the mean) was measured for the single-trial classifications with EEG feature vectors from the calibration phase, which was significantly better than the chance level of 0.5 ($Z = 3.37, p < 0.05$). Adding the fixation duration as extra feature did not improve the results, the AUC remained at the same level (significantly better than chance; $Z = 3.37, p < 0.05$). When only the fixation duration served as feature, an AUC of 0.51 ± 0.01 was obtained, which was not significantly better than chance ($Z = 1.05, p > 0.05$, Bonferroni corrected for the three Wilcoxon signed rank tests on the population level).

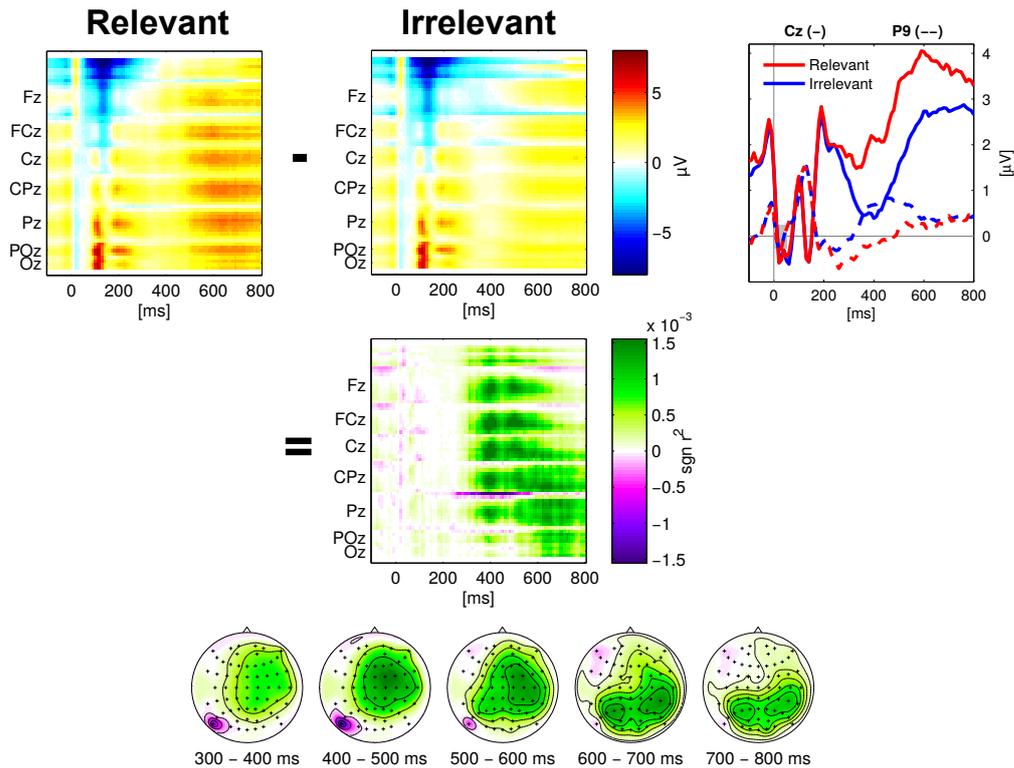


Figure 4.3: Patterns in the EEG differed when the word read was relevant to the category of interest or irrelevant (calibration phase). *Top*: EEG time series for relevant and irrelevant words (for all channels sorted from front to back and from left to right, and for two selected channels). *Centre*: Difference. *Bottom*: Topographies of the difference. [Figure from Wenzel et al., 2017, © IOP Publishing. Reproduced with permission. All rights reserved.]

Furthermore, the EEG patterns corresponding to relevant and irrelevant words were characterised in order to understand on which processes the classification success was based on (cf. figure 4.3). The EEG signal was inspected that followed the landing of the eye gaze on the words. The onset of the eye fixation was situated at $t = 0$ ms. Early components (until about 150 ms) were related to the saccade offset (respectively the fixation onset) and occurred equally in both conditions. Later components differed depending on whether the word was relevant or irrelevant. Relevant words evoked a left lateralised posterior negativity in comparison to irrelevant words and a positivity that shifted from fronto-central to parietal sites on both hemispheres. For this analysis, all EEG epochs of all participants were averaged separately for relevant and irrelevant words (cf. figure 4.3, top) and the difference between the two classes was assessed with signed squared biserial correlation coefficients (cf. figure 4.3, centre and bottom). Each time point measured at each EEG electrode was treated separately in order to characterise the spatio-temporal evolution. A significance threshold was not applied in order to show also subtle differences that can potentially be exploited by a multivariate classifier.

Relevant words were fixated for about $227.4 \text{ ms} \pm 8.7 \text{ ms}$ and irrelevant words for about $216.8 \text{ ms} \pm 7.8 \text{ ms}$ during calibration (mean \pm standard error of the mean). A paired t-test

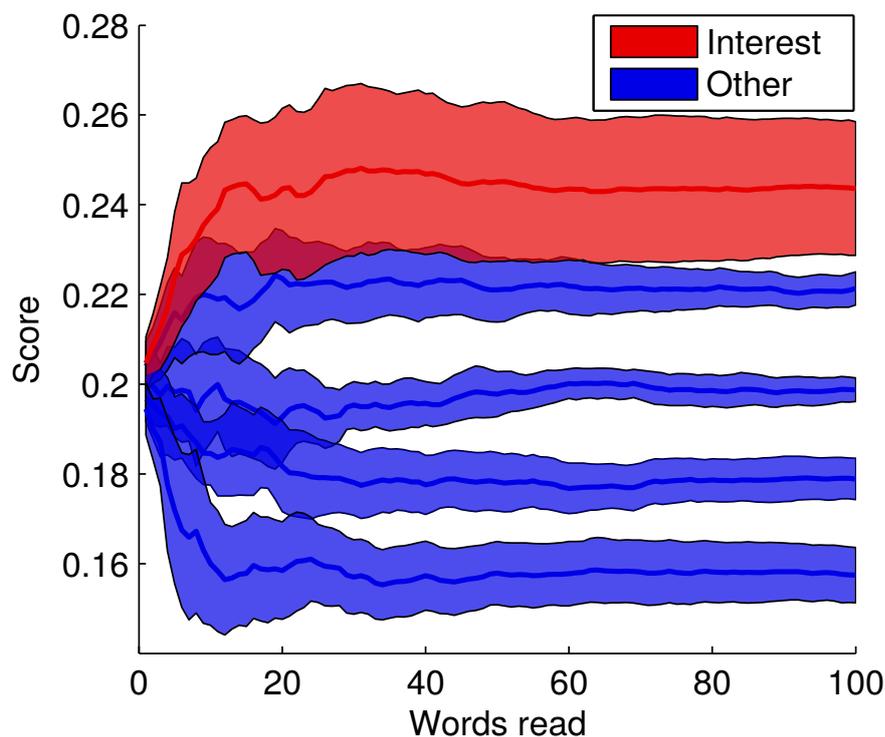


Figure 4.4: Evolution of the scores corresponding to the category of interest (red) and to the four other categories (blue, sorted according to the respective final score) during the online phase (combined EEG and gaze features). Tubes indicate the standard error of the mean. [Figure from Wenzel et al., 2017, © IOP Publishing. Reproduced with permission. All rights reserved.]

detected a significant difference between the two classes on the population level; $t(14) = 4.3, p < 0.05$.

4.3.2 Online prediction

The previously trained classifiers were applied during the online phase to the incoming data and it was predicted for each word if it was relevant to the category of interest or not. The class membership probability estimates were averaged per semantic category and the obtained five-dimensional vector was normalised to unit length (cf. section 4.2.2). Figure 4.4 displays the evolution of the resulting scores corresponding to the category of interest and to the four other categories, which were sorted according to the respective final score (combined EEG and gaze features; average over all participants). With more words being read by the participant, the score of the category of interest grew in comparison to the other categories. Note that the splitting of the four 'other' categories is a selection effect.

Figure 4.5 shows the evolution of the rank of the category of interest among the five semantic categories (combined EEG and gaze features; average over all participants). The category of

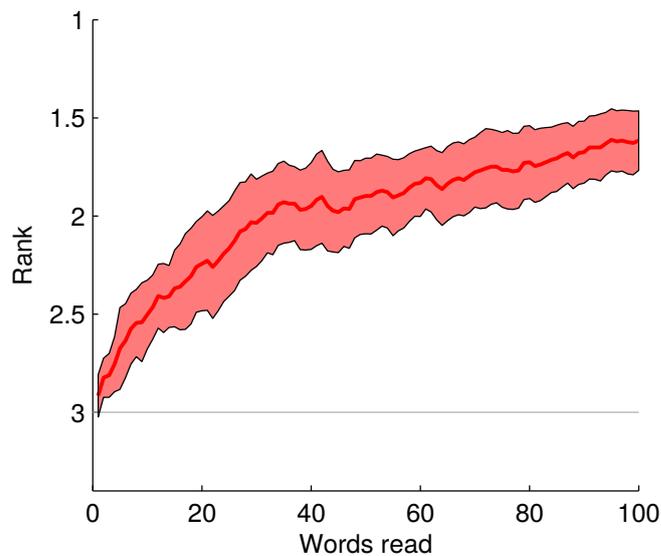


Figure 4.5: Evolution of the rank of the category of interest among the five categories during the online phase (combined EEG and gaze features; note the direction of the y-axis with the top rank of 1 on top; the shaded area indicates the standard error of the mean). [Figure from Wenzel et al., 2017, © IOP Publishing. Reproduced with permission. All rights reserved.]

interest started with an average rank of three and moved towards the top of the ranking with more words being read (note the direction of the y-axis).

Table 4.1 lists the average final rank of the category of interest for each single participant (i.e. when all hundred words per repetition had been read; cf. section 4.2.2). The predictions were based on feature vectors including either the EEG data or the fixation duration, or a combination of the two measurement modalities (columns in the table). The final rank was below three in every single participant when only EEG features were used and even smaller when the fixation duration was added as extra feature. Deploying the fixation duration as single feature resulted in a comparably large final rank. On the population level, the final rank was significantly below three for all feature types ($Z_{\text{EEG\&Gaze}} = -3.38$, $Z_{\text{EEG}} = -3.38$, $Z_{\text{Gaze}} = -2.41$, $p < 0.05$, Bonferroni corrected for the three Wilcoxon signed rank tests).

Figure 4.6 displays the EEG patterns during the online phase for relevant and irrelevant words. The spatio-temporal patterns evolve in the online phase (cf. figure 4.6) similar to the calibration phase (cf. figure 4.3) until about 500 ms. Relevant words evoked a posterior negativity and a central positivity in comparison to irrelevant words. In the online phase, a negativity on the left hemisphere started at 500 ms, in contrast to the calibration phase where the central positivity continued.

Relevant words were fixated for about $239.5 \text{ ms} \pm 12.4 \text{ ms}$ and irrelevant words for about $208.2 \text{ ms} \pm 7.0 \text{ ms}$ during the online phase (mean \pm standard error of the mean). The two classes differed significantly on the population level according to a paired t-test; $t(14) =$

Table 4.1: Final ranks of the online phase when respectively hundred words had been read (averages over the seventeen repetitions per participant, as well as over all participants). The combined and the single modalities are listed separately.

Participant	EEG & Gaze	EEG	Gaze
1	1.29	1.35	2.59
2	1.12	1.12	1.18
3	1.53	1.65	4.59
4	1.53	1.47	2.00
5	2.12	2.47	2.24
6	1.76	1.76	2.53
7	1.06	1.12	1.76
8	1.53	1.53	2.41
9	1.47	1.47	2.76
10	1.65	1.65	2.00
11	1.88	1.88	3.29
12	1.76	2.06	1.47
13	2.00	2.00	1.88
14	1.65	1.71	2.82
15	1.88	1.94	1.53
Mean \pm SEM	1.62 \pm 0.08	1.68 \pm 0.09	2.34 \pm 0.22

4.7, $p < 0.05$.

4.4 Discussion

4.4.1 Calibration

All participants complied with the task instructions because they recalled the words that were relevant to the selected semantic category with an accuracy of at least 72 % (giving random answers would result in an expected accuracy of about 20 % due to the five possible categories). EEG and eye tracking signals recorded during the calibration phase were used to train classifiers (individually for each participant) to discriminate relevant words from irrelevant words.

The trained EEG-based classifiers were able to generalise to unseen data, because the cross-validation results with calibration data were significantly better than it can be expected from random guessing (cf. section 4.3.1; please note that the AUC served as straightforward metric here, in contrast to the online phase where the ranking of the categories provided a more descriptive metric). Classification was apparently possible because relevant words evoked a different neural response than irrelevant words (cf. section 4.3.1 and figure 4.3). In previous research on brain-computer interfacing, the stimuli of interest evoked a similar neural response with a left lateralised negativity and a central positivity (cf. figure 2, right panel, in Treder et al., 2011), even though the stimuli used in the cited study were not words but geometric shapes

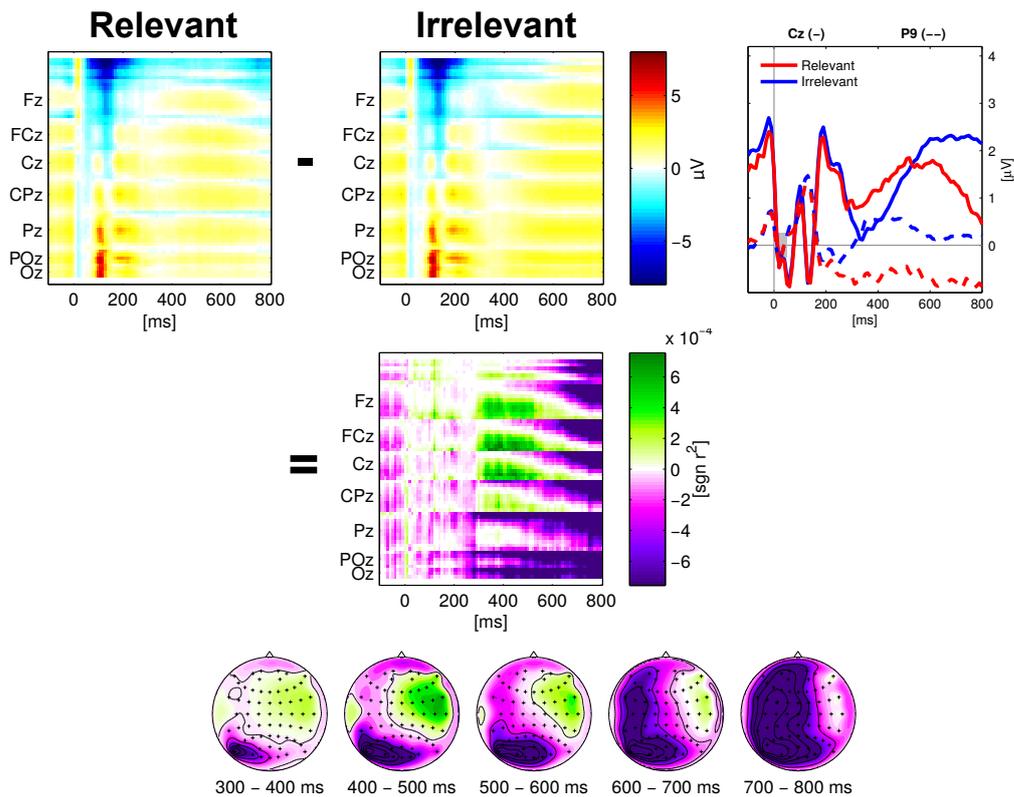


Figure 4.6: EEG patterns during the online phase. *Top*: EEG time series for relevant and irrelevant words (for all and for two selected channels). *Centre*: Difference. *Bottom*: Topographies of the difference. [Figure from Wenzel et al., 2017, © IOP Publishing. Reproduced with permission. All rights reserved.]

flashed on the screen while the eyes did not move. Hence, it was shown with the present investigation that the methods developed for brain-computer interfacing can be employed for inferring the relevance of words under unrestricted viewing conditions.

Concatenating the fixation duration to the feature vectors did not improve the predictive performance. Relevant words could not be detected in single-trial, when the fixation duration served as only feature for the classifications (using data from the calibration phase). Nevertheless, a small but significant difference of the fixation duration between the two classes was found on average (cf. section 4.3.1).

4.4.2 Online prediction

It was predicted in real-time which words were relevant for the reader, who was looking for words related to a semantic category of interest. The five categories were ranked according to the normalised five-dimensional average score vector. Perfect prediction of the category of interest would have resulted in a score of 1 and a rank of 1 for the category of interest. If each word was classified randomly as relevant or irrelevant, an average score of 0.2 and an average

rank of 3 can be expected. The score and the rank of the category of interest started at this chance level, as it can be assumed. With more words being read, the score grew and the rank decreased (cf. figures 4.4 and 4.5). Apparently, evidence could be accumulated by integrating information over the incoming single predictions.

The combination of EEG and fixation duration resulted in the best predictive performance (cf. table 4.1). The gaze did not contribute much to the relevance estimate because features from the EEG alone were more informative than when the fixation duration was used as single feature (while it has to be considered that information about the eye gaze is required for the EEG feature extraction, because the EEG signals had to be related to the corresponding words looked at; cf. section 4.2.1).

The successful transfer of the classifiers from the calibration phase to the online phase is reflected in the underlying data. The EEG patterns evolved similarly in the calibration and in the online phase up to about 500 ms after fixation onset (cf. figures 4.3 and 4.6). The later discrepancy is presumably a result of the different tasks, because the relevant words had to be memorised only in the calibration phase (cf. sections 4.2.1 and 4.2.2).

4.4.3 Conclusion

The study demonstrates that the subjective relevance of words for a reader can be inferred from EEG and eye gaze in real-time. The methods employed are rooted in research on brain-computer interfacing based on event-related potentials, where stimulus recognition is usually sufficient, and where sequences of single stimuli are typically flashed. In contrast, the investigation presented here is characterised by the requirement to interpret words with respect to their semantics, and by the presentation of several words at the same time. Neural activity was related with eye tracking to the respective word read. The typically employed counting task was avoided because it would not be sensible for *implicit* relevance detection [cf. Wenzel et al., 2016b]. The task instruction during the online phase was merely to look for (and not to count) words relevant to the category of interest. Task engagement was additionally fostered by explaining the predictive mechanism underlying the adaptive visualisation. The experiment exploits a situation that allows for integrating implicit information across several single words. In a next step, the methods could be applied to a situation where sentences or entire texts are being read, which will entail a number of new challenges for the data analysis. While this study serves as a proof-of-principle, the methods can potentially be used in the future for mapping the subjective relevance of the field of view in different applications (cf. section 4.1). In summary, this study represents a further step towards inferring the interest of a person from information implicitly contained in neurophysiological signals.

4.5 Lessons learned

- By combining information from neural activity and eye movements, it was decoded online which words were subjectively relevant for a reader
- For this purpose, high dimensional feature vectors were extracted from the recorded EEG and eye tracking signals in real-time, and were classified with regularized linear discriminant analysis.
- Subjective relevance maps of the visual surrounding can give insights into the interest of a person in real-time.
- Limitations: The relevance estimates come with a considerable uncertainty, and evidence had to be accumulated over time. The words read were not syntactically interrelated.

5 Variable salience challenges the inference from EEG and eye gaze

5.1 Introduction

The salience of the items present in the field of view has to be considered for the inference of relevance maps from EEG and eye gaze for the reasons set out below. Single stimuli are typically flashed in sequence in brain-computer interfacing. Therefore, the timing of stimulus recognition is precisely known. In contrast, our regular visual environment consists of various elements present at the same time (exemplified in figure 5.1). For instance, computer users operating a web browser view numerous words and pictures displayed side by side on the screen. The single items are not flashed one by one but are fixated sequentially with the eye gaze. For this reason, eye tracking is required for estimating if an item was either relevant or irrelevant to the objectives of a person. First, it is necessary to know the position of the eye gaze for relating the ongoing neural activity to the respective item looked at. Second, BCIs based on event-related potentials require time markers of reference for extracting features from the continuous EEG signal [cf. Blankertz et al., 2011, and chapter 2.3.3.1]. The eye gaze jumps with saccades from one fixated position to the next, which can serve as time marker of reference (cf. chapter 4).

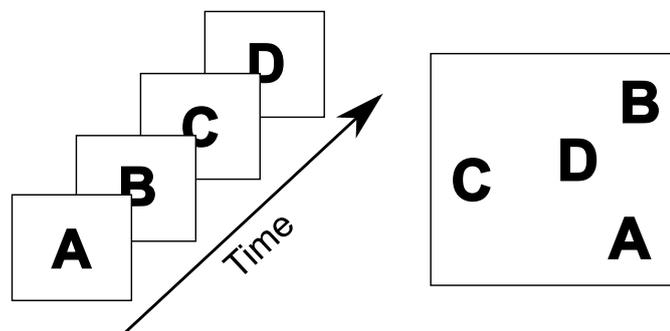


Figure 5.1: Single stimuli are typically flashed in sequence in ERP-based BCIs (left). The regular visual environment consists of various elements present at the same time (exemplified on the right). The measured neural activity can be related with eye tracking to an item looked at.

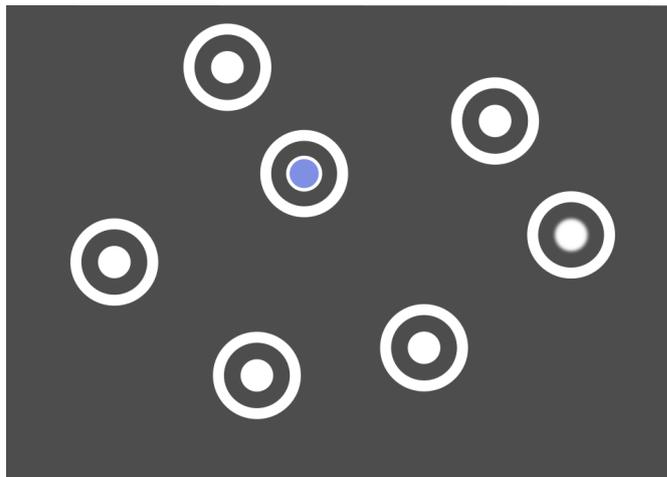


Figure 5.2: The example illustrates that the timing of recognition depends on the salience. The discs represent elements of the visual surrounding (e.g. pictures or words shown on a computer screen) and have a mixed salience of discriminative information. The blue disc clearly stands out (high salience) in comparison to all other discs. Yet, one of the white discs differs also from all other discs, which becomes *apparent only at the second glance* (low salience; the white disc on the right is blurred).

However, the visual environment is diverse and can contain items of a mixed salience of discriminative information (exemplified in figure 5.2). Light entering the eye along the line of sight falls onto the fovea where the retina has the highest visual acuity. Peripheral retinal areas provide a lower spatial resolution [Wandell, 1995]. Accordingly, items of a low salience may be recognised only after the landing of the eye gaze on the item, which is now captured in high-resolution by the fovea. In contrast, highly salient items can be recognised already in peripheral vision, i.e. before or even without a shift of the gaze towards the item. As a result, people can classify eye-catching items immediately as relevant or irrelevant, but inconspicuous items only at a later time point. For this reason, the timing of the corresponding neural processes varies with respect to the saccades that serve as time markers of reference. The resulting jitter can be problematic for the intended relevance detection, because state-of-the-art BCI methods were developed for sequences of flashed stimuli and assume that the discriminative EEG activity is tightly time locked to the reference time point.

Hence, it was tested if relevance maps can be inferred from EEG and eye gaze even when the salience of the items present in the field of view is mixed. The participants of an experimental study performed a search task where the salience of the target items was varied. EEG and eye movements were recorded and it was estimated from the signals which items attracted the particular attention because they were targets and thus task relevant. Simple geometrical shapes represented the variable components of our real visual environment. The shapes were designed with a pronounced variation of salience in order to achieve a large jitter of recognition with respect to the eye movements (cf. figure 5.2 and section 5.2.1).

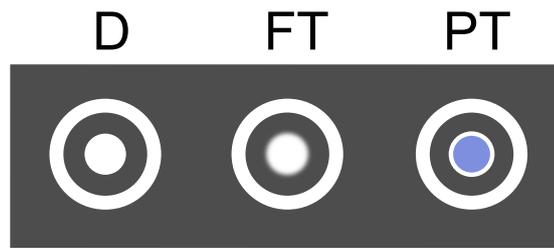


Figure 5.3: Distractor (D), foveal target (FT) and peripheral target (PT). [Figure from Wenzel et al., 2016c, reproduced with permission.]

The results of the study show that the BCI-based detection of the stimuli of interest is also possible when the stimulus salience is mixed, and when eye movements serve as time markers of reference. The methods were apparently able to cope with the variable timing of the neural activity. Information contained in EEG and eye tracking data was found to be complementary and neural signals were captured despite of the unrestricted eye movements that can potentially interfere with the EEG signals. In summary, it was demonstrated how EEG and eye tracking data can provide implicit information about the relevance of items that feature a pronounced variation of salience and that are displayed side by side at the same time on a screen.

This chapter is based on the following paper:

Wenzel, M. A., Golenia, J.-E., and Blankertz, B. (2016c). Classification of eye fixation related potentials for variable stimulus saliency. *Frontiers in Neuroprosthetics*, 10(23). doi: 10.3389/fnins.2016.00023.

5.2 Materials and methods

5.2.1 Experimental design

The sixteen participants of the study performed a gaze contingent search task while the electroencephalogram was recorded and the eye movements were tracked. Twenty-four items situated at random positions on the screen had to be scanned and the number of targets among the distractors had to be reported. The salience of target discriminative information was varied by using two types of targets, which could be either recognised already in peripheral vision or only in foveal vision. Peripheral targets (PT) featured a blue disc and could be discriminated from the white discs of the distractors (D) already in peripheral vision (cf. figure 5.3). In contrast, foveal targets (FT) featured a white *blurred* disc and could be discriminated from the similar distractors only in foveal vision. Accordingly, they had to be fixated for target detection (cf. section 5.4.4).

Fixations were not necessary for target detection (cf. section 5.4.4) but were nevertheless required for task accomplishment (cf. last paragraph in this section 5.2.1). The eye movements

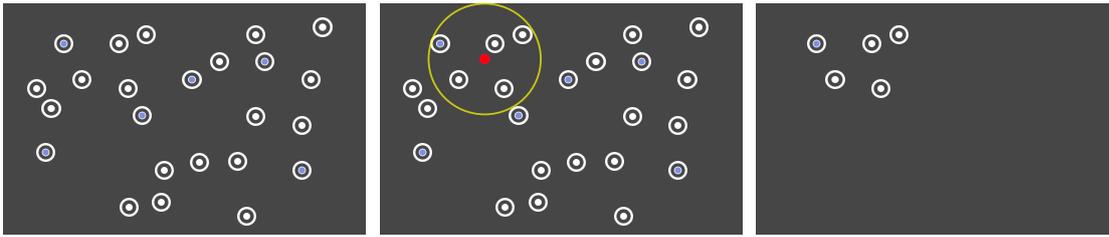


Figure 5.4: Illustration of the gaze contingent search task. The current point-of-gaze controlled the disclosure of the items. *Left*: The arrangement of the items was predefined. *Center and right*: Only items within a certain radius (yellow) around the current point-of-gaze (red) were dynamically disclosed. [Figure from Wenzel et al., 2016c, reproduced with permission.]

were *not* restrained (e.g. by demanding slow movements). Foveal targets and distractors were presented in the experimental condition **F** and peripheral targets and distractors in the condition **P**. Both types of targets and distractors were shown in the mixed condition **M**, which modelled the variation of salience that can be expected in realistic settings.

The dice were rolled for each of the 24 items displayed on the screen to decide if it is a target or a distractor (repeated for every repetition of the search task). Each item had the independent chance of being a target with a probability of 25% (allocated to 12.5% foveal and 12.5% peripheral targets in the mixed condition **M**). On average, there were 5.9 ± 2.2 (mean \pm std) targets presented ranging from 1 to 12. The layout of the 24 items was predefined for each repetition of the search task (cf. last paragraph of this section). The items were initially hidden and were disclosed area by area, depending on the eye gaze (cf. figure 5.4). All items within a radius of 250 pixels (visual angle of 6.7°) around the current point-of-gaze were uncovered. When moving the gaze, previously hidden items appeared at the boundary of this circle. Thus, all items appeared in peripheral vision. The gaze contingent stimulus presentation was updated with 30 Hz based on the continuous eye tracker signal sampled with 250 Hz. After leaving the radius of 250 pixel, the items disappeared again 1.5 seconds later. This gaze contingent disclosure impeded the detection of all peripheral target items more or less at once by an unfocused ‘global’ view on the whole screen. Moreover, it allowed for studying the neural response to the stimulus appearance in peripheral vision.

Every item could disappear and reappear again in the gaze contingent stimulus presentation. However, as soon as an item was directly fixated (detected by the online algorithm of the eye tracker), it disappeared 1.5 seconds later and did not reappear again. Note that it was not necessary to fixate the item for 1.5 seconds. This behaviour forced the participants to discriminate between targets and distractors upon the first fixation of an item and impeded the careless gaze on items, which would presumably attenuate components of the event-related potential (ERP) that are related to target recognition.

The three conditions of the search task were repeated 100 times each resulting in 300 repetitions in total. Before the beginning of each repetition of the search task, a fixation cross had to be fixated until it disappeared after two seconds. As soon as all target items had been fixated,

the stimulus presentation ended and the question was asked to enter the number of targets. Finally, the participant was informed if the answer was correct or not by a “happy” or a “sad” picture to enhance task engagement. Ten subsequent repetitions of one condition built one block. The blocks of the three conditions were interleaved and the participants were informed about the respective condition at the beginning of each block.

5.2.2 Experimental setup

The participants were seated in front of a screen at a viewing distance of sixty centimetres and entered the counted number of targets with a computer keyboard. An eye tracker (*RED 250*, SensoMotoric Instruments, Teltow, Germany; sampling frequency of 250 Hz) was attached to the screen and a chin rest gave orientation for a stable position of the head. The gaze contingent stimulus presentation was updated with 30 Hz. The screen itself had a refresh rate of 60 Hz, a resolution of 1680 x 1050 pixels, a size of 47.2 cm x 29.6 cm and subtended a visual angle of 38.2° in horizontal and of 26.3° in vertical direction. The target and distractor items had a diameter of 50 pixels, subtended a visual angle of 1.3°, and had a minimal distance of 70 pixels or 1.9° between each other and of 100 pixels or 2.7° from the border of the screen. An item was considered as fixated if the fixation position was situated within a radius of 75 pixels or 2.0° from the centre of the item and no other item was closer.

EEG signals were recorded with 64 active electrodes (*BrainAmp, ActiCap*, BrainProducts, Munich, Germany; sampling frequency of 1000 Hz). The ground electrode was placed on the forehead, the reference electrode on the left mastoid and one electrode on the right mastoid for later re-referencing (see section 5.2.3). One of the electrodes was stuck below the left eye for electrooculography. The vertical electrooculogram (EOG) was computed by subtracting the electrode Fp1 from the electrode below the left eye. The horizontal EOG was yielded by subtracting the electrode F9 from the electrode F10.

To accomplish the dynamic stimulus presentation and multimodal data acquisition, *Matlab* and *Python* code was written. The following software programs were running on two computers and interacting: *Pyff* for stimulus presentation [Venthur et al., 2010], *BrainVisionRecorder* (BrainProducts, Munich, Germany) for EEG data acquisition, *iView X* (SensoMotoric Instruments, Teltow, Germany) for eye tracking and online fixation detection and the *iView X API* to allow for communication between the computers (see figure 5.5 for a schematic representation).

5.2.3 Data acquisition

Sixteen persons with normal or corrected to normal vision and no report of eye or neurological diseases participated in the experiments. The age of the four women and twelve men ranged from 18 to 54 years and was on average 30.7 years. One recording session included giving an informed written consent to take part in the study, vision tests for visual acuity and eye

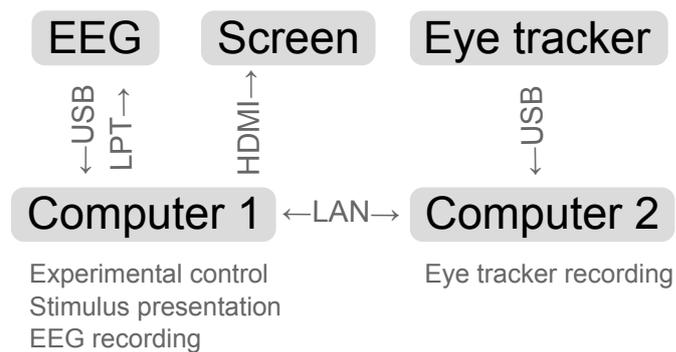


Figure 5.5: Schematic representation of the experimental setup. Arrows indicate the data flow between the devices. [Figure from Wenzel et al., 2016c, reproduced with permission.]

dominance, preparation of the sensors, eye tracker calibration and validation, introduction to the task and to the gaze contingent stimulus presentation, training runs, the main experiment (with a duration of about two hours) and standard EEG measurements [eyes-open/closed, simple oddball paradigm, see Duncan et al., 2009]. The proper calibration of the eye tracker was re-validated and – if necessary – re-calibrated in the middle of the experiment and in the case that the subject reported that the items did not disappear after fixation. The study was approved by the ethics committee of the Department of Psychology and Ergonomics of the Technische Universität Berlin (reference BL_01_20140120).

The start times of the first fixations of targets and distractors were determined from the eye tracker signal sampled at 250 Hz with the software of the eye tracker (*IDF Event Detector*, SensoMotoric Instruments, Teltow, Germany; event detection: ‘high speed’, peak velocity threshold: 40°/s, min. fixation duration: 50 ms). The synchronously recorded EEG and eye tracking signals were aligned with the help of the sync triggers, which had been send via parallel port interface (LPT) to the EEG system every second during the experiment, and the time-stamps of the eye tracker logged at the same time. The parameters of the function that mapped eye-tracking-time to EEG-time were determined with linear regression. The EEG data were low-pass filtered with a second order Chebyshev filter (42 Hz passband, 49 Hz stopband), down-sampled to 100 Hz, re-referenced to the linked-mastoids and high-pass filtered with a finite impulse response filter at 0.1 Hz.

5.2.4 Data analysis

5.2.4.1 Compliance check

The compliance of the participants with the instructions was checked by assessing the performance in the search task. For this purpose, the percentage of correct responses was computed as well as the absolute differences between response and true number of targets. It was tested whether the experimental conditions differed in these respects with one-way repeated measures analyses of variance.

5.2.4.2 Target estimation with EEG and eye tracking features

Based on EEG and eye tracking data, it was estimated which items displayed on the screen were targets, and accordingly relevant for the person to solve the search task, and which were distractors. For this purpose, feature vectors were extracted from EEG and eye tracking signals, labelled either as *target* or as *distractor* depending on the corresponding item, and classified.

EEG feature extraction. The continuous multichannel EEG time-series were segmented in epochs of 0 ms to 800 ms relative to the onset of the first fixation of each item. Each epoch was channel-wise baseline corrected by subtracting the mean signal within the 200 ms before the fixation-onset. The EEG signal measured at each channel was then averaged over 50 ms long intervals and the resulting mean values of all channels and all intervals were concatenated in one feature vector per epoch (that represents the spatio-temporal evolution of the neural processes, as observed at the electrodes). Improved classification performance is intended goal of this step – via a reduction of the dimensionality of the feature vectors in comparison to the number of samples [cf. the section ‘Features of ERP classification’ in Blankertz et al., 2011].

Eye tracking feature extraction. From the eye tracking data, the duration of the first fixation of each item and the duration and distance of the respective previous and following saccade were determined and used as features.

Classifications. EEG and eye tracking features were classified both separately (‘EEG’, ‘ET’) and together (‘EEG & ET’) – by appending the eye tracking features to the corresponding EEG feature vectors – with regularized linear discriminant analysis. The shrinkage parameter was calculated with an analytic method [see Friedman, 1989; Ledoit and Wolf, 2004; Schäfer and Strimmer, 2005, for more details]. More information about this approach to single-trial ERP classification is provided in Blankertz et al. [2011]. The classification performance was evaluated in 10x10-fold cross-validations with the area under the curve (AUC) of the receiver operating characteristic, which is applicable for imbalanced data sets [more distractors than targets; Fawcett, 2006]. The better the classification performance, the more the AUC differs from 0.5. The classifications were performed separately for each combination of participant, experimental condition (F, P, M) and modality (‘EEG’, ‘ET’, ‘EEG & ET’). Per condition and modality, it was assessed with one-tailed Wilcoxon signed-rank tests whether the median classification performance of all participants was significantly better than the chance level of an AUC of 0.5.

Electrooculogram. The classifications were additionally performed using only the horizontal and the vertical electrooculogram (‘EOG’). The same feature extraction method was employed for the EOG channels as for the EEG. The aim was not to get the best possible classification from the EOG, but to check whether the performance of the EEG-based classification is in part based on EOG signals and, therefore, can be explained to a certain extent by eye movements as confounding factor.

Subsequently, it was tested with a two-way repeated measures analysis of variance, if the

experimental conditions (F, P, M) and the modalities ('EEG', 'ET', 'EEG & ET', 'EOG') had an effect on the classification performance.

Two additional analyses of the EEG data of the mixed condition M were conducted, where there were both peripheral and foveal targets present as well as distractors:

- A *combined classifier* consisting of a combination of two classifiers was designed. One classifier was trained to discriminate foveal targets from distractors and another classifier learned to discriminate peripheral targets from distractors – both using fixation-aligned EEG epochs from condition M. The two classifiers were then applied to the respective test-subset of a 10x10 crossvalidation, where the salience of the target items (foveal or peripheral) was not unveiled. The posterior probabilities yielded from the two classifiers were averaged for each EEG epoch to predict if it was a target or a distractor epoch [Tulyakov et al., 2008]. It was tested if the combined classifier was better able than the standard classifier to cope with the temporal variability of the neural response in relation to the eye movements, which was present in the mixed condition M, and, which can be expected in realistic settings.
- A reference case for the achievable classification performance would be represented by a *split analysis*, where peripheral and foveal targets are treated separately. This models a situation (which usually can not be expected in the application case) where the salience of each item is known and, accordingly, a situation where it is known whether the item can be recognised in peripheral vision or not. For this purpose, the EEG data of the mixed condition M were split and either foveal or peripheral targets were classified against distractors using fixation-aligned EEG epochs. The distractor data were split arbitrarily in halves.

Appearance-aligned EEG features. Furthermore, it was tested if information was present in the EEG data already when the items appeared in peripheral vision, i.e. even before fixation-onset (cf. the description of the gaze-contingent stimulus presentation in section 5.2.1). For this purpose, the EEG time-series were segmented in epochs aligned to the first appearance of each item on the screen. Baseline correction of the 800 ms long epochs was performed using the 200 ms interval before the appearance. Features were extracted and classified as described above for the fixation-aligned EEG epochs.

5.2.4.3 Characteristics of target and distractor EEG epochs

The EEG data were further characterised to provide insights into the underlying reasons for success or failure of the classifications and into the neural correlates of peripheral and foveal target recognition.

EEG epochs aligned to item appearance and fixation. The EEG time-series were segmented in epochs aligned to the first appearance of each item on the screen (caused by gaze movements,

cf. section 5.2.1) and in epochs aligned to the first fixations of the items (cf. section 5.2.4.2). Each 1000 ms long epoch started 200 ms before the appearance or fixation, was channel-wise baseline corrected by subtracting the mean signal within the 200 ms interval before the respective event and was labelled as *target* if the corresponding item was a target and otherwise as *distractor*.

Class-wise averages. Single EEG epochs contain a superposition of different components of brain activity, including non-phase locked oscillatory signals. Averaging the EEG epochs attenuates the non-phase locked components. The average is referred to as the event-related potential, which is abbreviated as ERP. To single out the phase locked brain activity, target and distractor EEG epochs of all participants were class-wise averaged. The two types of events (appearance, fixation-onset) and the three experimental conditions (E, P and M) were assessed separately. Before averaging, artefacts were rejected with a heuristic: channels with a comparably small variance were removed as well as epochs with a comparably large variance or with an absolute signal amplitude difference that exceeded $150 \mu\text{V}$ (only the interval of 800 ms after the appearance or fixation was considered for artefact rejection). Artefact rejection was used for the visualization in order to obtain clean event-related potentials. The challenge was taken on for single-trial classification of dealing with trials that are corrupted by artefacts as this is beneficial for online operation in future use cases. Due to the usage of data-driven multivariate methods, many types of artefacts can indeed be successfully projected out. The influence of eye movements on the EEG data are discussed in the sections 5.4.2 and 5.4.5.

Statistical differences between classes. Target and distractor EEG epochs were compared with univariate statistics. Differences between the epochs of the two classes were quantified per subject, for each channel, and each time point with the signed squared biserial correlation coefficient (signed r^2) between each univariate feature and the class label (+1 for targets and -1 for distractors). A signed r^2 of zero indicates that feature and class label are not correlated and a positive value indicates that the feature was larger for targets than for distractors and vice versa. In an across-subject analysis, the individual coefficients were aggregated into one grand average value for each univariate feature. The p-value related to the null hypothesis that the signed r^2 across all subjects is zero was derived.

Classifications with either spatial or temporal EEG features. While spatio-temporal EEG features served for the actual classification purpose (cf. section 5.2.4.2), the classification with either temporal features or spatial EEG features made it possible to specify where the discriminative information resided in space and time [see Blankertz et al., 2011]. In the case of temporal features, the time-series were classified separately for each EEG channel, using the interval of 800 ms post-event. The AUC-scores obtained for each channel were averaged over participants and displayed as scalp maps. In the case of spatial features, the EEG epochs were split in 50 ms long (multichannel) chunks, which were averaged along time. The resulting feature vectors were classified separately for each chunk and the mean AUC-scores of all participants were displayed as time courses.

5.2.4.4 Eye gaze characteristics

The eye movements of the participants were characterised with the average fixation duration of each item type in each experimental condition. In addition, the fixation frequency was computed, i.e. the number of the fixations on each item type in comparison to the total number of fixations on all item types. Besides, the average duration and distance of the first saccades to the items and of the respective following saccades were calculated. Moreover, the average latency between the first appearance of each item and its fixation were determined. Re-fixations of items were not considered because, then, the identity of the item had been already revealed.

5.3 Results

5.3.1 Compliance check

The participants gave correct responses in condition F in 70.6 %, in P in 81.3 % and in M in 75.1 % of the cases. The absolute differences between response and true number of targets were 0.370 in F, 0.255 in P and 0.350 in M. These two performance measures differed significantly between conditions (one-way repeated measures analyses of variance, $F(2, 30) = 11.7$, $p \leq 0.01$ and $F(2, 30) = 5.88$, $p \leq 0.01$).

5.3.2 Target estimation with EEG and eye tracking features

It was estimated which items displayed on the screen were targets of the search task based on EEG and eye tracking data. The results of the classifications are listed in table 5.1. The two modalities were either classified together ('EEG & ET') or separately ('EEG', 'ET'). Additionally, features only from the electrooculogram ('EOG') were used to investigate to which degree eye movements might have confounded the classifications with EEG features. The classification performance was better than chance in all experimental conditions and for all modalities except for the EOG features (one-tailed Wilcoxon signed-rank tests, $p \leq 0.01$, Bonferroni corrected for the twelve comparisons).

The modalities as well as the experimental conditions had a significant effect on the classification performance (two-way repeated measures analysis of variance, $F(3, 165) = 203$, $p \leq 0.01$ and $F(2, 165) = 14.6$, $p \leq 0.01$).

Using EEG and eye tracking features in combination resulted in classification performances that were significantly better than when either eye tracking or EEG features were used alone. Significantly better results were obtained with eye tracking features than with EEG features (one-tailed Wilcoxon signed-rank tests, $p \leq 0.01$, Bonferroni corrected for the three comparisons). The individual classification performances ranged from 0.556 to 0.828 in the case of 'EEG & ET', from 0.529 to 0.765 in the case of 'EEG', and from 0.543 to 0.862 in the case of 'ET' (averages and standard deviations are listed per condition in table 5.1). The individual results

Table 5.1: Classification results are listed for the different modalities and the three experimental conditions. ‘EEG & ET’ denotes the multimodal classification of EEG and eye tracking features, ‘EEG’ and ‘ET’ stand for the separate assessments with either one or the other modality and ‘EOG’ for the classifications with features from the electrooculogram only. In the table, AUC-scores are presented as averages over participants together with the corresponding standard deviations. Asterisks mark results significantly above the chance level 0.5 (one-tailed Wilcoxon signed-rank tests, $p \leq 0.01$, Bonferroni corrected for the twelve comparisons).

	F	P	M
EEG & ET	0.726* \pm 0.054	0.714* \pm 0.070	0.678* \pm 0.060
EEG	0.672* \pm 0.060	0.633* \pm 0.055	0.620* \pm 0.047
ET	0.677* \pm 0.044	0.718* \pm 0.084	0.652* \pm 0.061
EOG	0.516 \pm 0.020	0.536 \pm 0.033	0.514 \pm 0.020

for ‘EEG’ and ‘ET’ did not correlate significantly ($p > 0.01$).

The ranking of the three experimental conditions according to the classification performance was $F > P > M$ in the case of ‘EEG & ET’ and ‘EEG’ and $P > F > M$ in the case of ‘ET’ (cf. table 5.1). The classification performance was significantly better in condition F than in condition M in the cases of ‘EEG & ET’ and of ‘EEG’ and significantly better in P than in M in the case of ‘ET’ (one-tailed Wilcoxon signed-rank tests, $p \leq 0.01$, Bonferroni corrected for the three comparisons).

Per participant, condition and modality, about 544 target versus 1181 distractor samples were, on average, available for the classification. These numbers result from the about $24 * 0.25 * 100 = 600$ targets and $24 * 0.75 * 100 = 1800$ distractors presented in total and the fact that not all items were fixated by the participant (cf. section 5.3.4 and table 5.3).

The results of the two additional analyses of the fixation-aligned EEG epochs from condition M are listed in table 5.2. For the *combined classifier*, one classifier had been trained to discriminate foveal targets from distractors and a second classifier to discriminate peripheral targets from distractors. Both classifiers were applied to the test data in combination by averaging the posterior probabilities yielded per epoch. The combined classifier performed, on average, slightly better than the standard EEG-based classifier (cf. table 5.1, row ‘EEG’, column ‘M’), however not significantly ($p > 0.01$). In the *split analysis*, either foveal or peripheral targets were classified against distractors. The performance of the classification of foveal targets versus distractors (‘FT vs. D’) was significantly better than the standard EEG-based classification of condition M ($p \leq 0.01$) and comparable to the result of condition F (cf. table 5.1, row ‘EEG’, columns ‘M’ and ‘F’).

Appearance-aligned EEG features. Classification performance was better in condition P than in the conditions F and M (F: 0.518 ± 0.018 , P: 0.637 ± 0.044 , M: 0.547 ± 0.023). In all conditions, the performance was significantly better than the chance level (one-tailed Wilcoxon signed-rank tests, $p \leq 0.01$, Bonferroni corrected for the three comparisons).

Table 5.2: Results of the combined classifier and the split analysis in condition M (averages and standard deviations of the sixteen participants of the study). All classification results were significantly above the chance level of an AUC of 0.5 (one-tailed Wilcoxon signed-rank tests, $p \leq 0.01$, Bonferroni corrected for the three comparisons). For the split analyses, the distractor data were split arbitrarily in halves (denoted as $D_{1/2}$ and $D_{2/2}$).

Method	Classes	[AUC]
Combined classifier	PT, FT vs. D	0.627 ± 0.042
Split analysis	FT vs. $D_{1/2}$	0.666 ± 0.065
Split analysis	PT vs. $D_{2/2}$	0.590 ± 0.036

5.3.3 Characteristics of target and distractor EEG epochs

5.3.3.1 Class-wise averages

The class-wise averages of the EEG epochs are presented in figure 5.6. The two types of events (*appearance* of an item on the screen and onset of the eye *fixation*, cf. section 5.2.4.3) and the three experimental conditions (F, P, M) were assessed separately. Electrode Pz was chosen for the presentation as time course, because it is well suited to capture the P300 wave [Picton, 1992]. Note that information regarding all electrodes and all time points is presented in the next section 5.3.3.2 with figure 5.7.

5.3.3.2 Statistical differences between classes

The statistical differences between target and distractor EEG epochs aligned either to item appearance or fixation are shown in figure 5.7. Significant differences ($p \leq 0.01$, Bonferroni correction for multiple comparisons due to the number of channels, time-points, conditions and event types) occurred mainly at central, parietal and occipital electrodes close to the midline of the head. Across-subject signed r^2 values that were not significantly different from zero were set to zero and remain white in the figure.

5.3.3.3 Classifications with either spatial or temporal EEG features

The results of the classifications of target versus distractor EEG epochs, using either spatial or temporal features are presented in the figures 5.8 and 5.9 respectively. The EEG epochs were aligned either to item appearance or fixation. The three experimental conditions F, P and M were assessed separately.

Spatial EEG features (i.e. data from separate time-intervals at all channels) of target versus distractor epochs were classified to characterise where the information resided in time. Figure 5.8 depicts the time courses of the classification performance averaged over subjects. In condition P, classification performance started to surpass the chance level of an AUC of 0.5 at about 200 ms after item appearance and reached the maximum at about 500 ms post-appearance

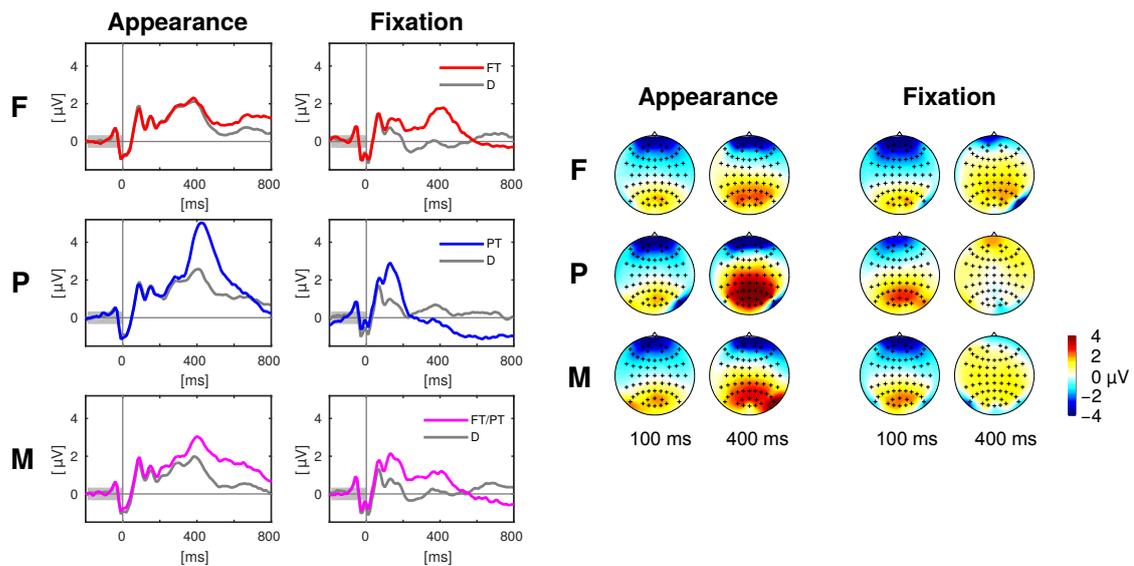


Figure 5.6: *Left*: Event related potentials aligned to appearance and fixation of targets (coloured) and distractors (gray) at the exemplary electrode Pz in the experimental conditions F, P and M. *Right*: The scalp maps depict the head from above with the nose on top and show the potentials averaged over 50 ms long intervals centred at 100 ms and 400 ms after the target appearance (left) and fixation (right). Please note that the positivity (‘yellow/orange/red’) at central, parietal and occipital electrodes was discriminative between targets and distractors in contrast to the negativity (‘blue’) at prefrontal and anterior frontal electrodes (cf. figure 5.7). Every figure throughout this section summarises the data of all sixteen participants of the study. [Figure from Wenzel et al., 2016c, reproduced with permission.]

with an AUC of about 0.6. In the conditions F and M, only a slight increase over time was observed after item appearance. In contrast, classification performance increased clearly in all three conditions in response to the fixation-onset. The maximum was reached faster in condition P, at about 150 ms, than in the conditions F and M, at about 300 ms. In condition P, the AUC values exceeded the chance level even before fixation-onset.

Temporal EEG features (i.e. the entire time-series of separate channels) were used to classify between target and distractor epochs to learn where the discriminative information resided in space. Figure 5.9 depicts the classification results as scalp maps (AUC-scores for each channel, averaged over participants). Channels situated at central, parietal and occipital positions showed the largest AUC-values and were, accordingly, most informative about the class membership.

5.3.4 Eye gaze characteristics

The fixation durations of the two or respectively three types of items differed significantly from each other in all experimental conditions (cf. figure 5.10; one-way repeated measures analyses of variance; F: $F(1, 15) = 28.9$, P: $F(1, 15) = 14.6$, M: $F(2, 30) = 17.3$; $p \leq 0.01$ respectively,

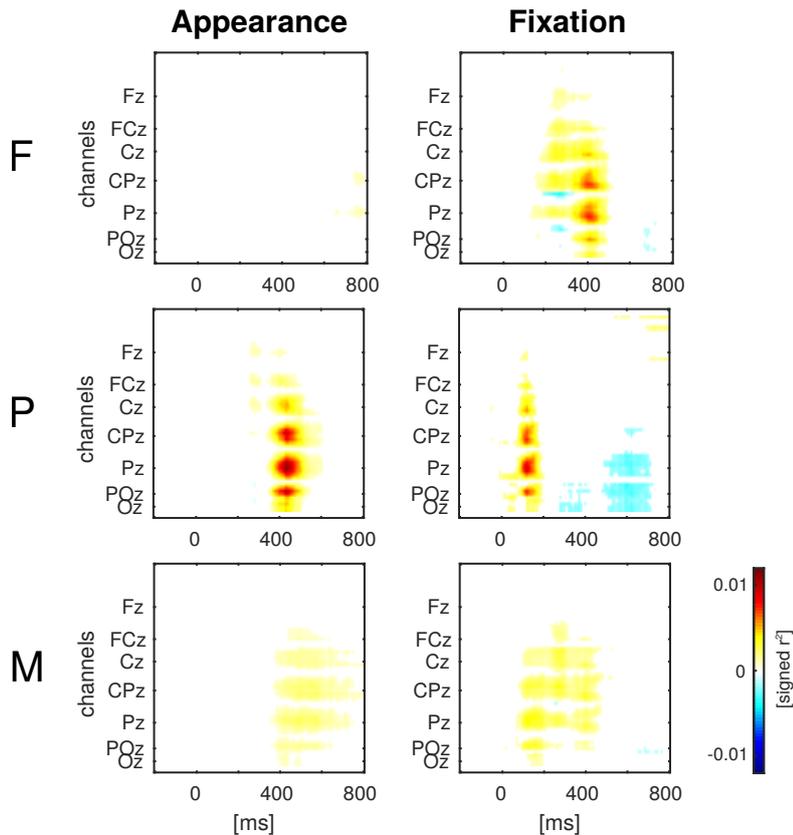


Figure 5.7: Statistical differences between target and distractor EEG epochs aligned to item appearance and fixation in the conditions F, P and M (across-subject signed r^2 values). The channels are ordered from the front to the back of the head (top to bottom in the figure). [Figure from Wenzel et al., 2016c, reproduced with permission.]

Bonferroni corrected for the three comparisons). On average, distractor items (D) were fixated shorter than target items (PT and FT).

The items were dynamically disclosed on the screen and could be subsequently fixated (cf. section 5.2.1). The latencies between the first appearances and the first fixations of the two or respectively three types of items differed significantly in all conditions (cf. figure 5.11, one-way repeated measures analyses of variance, condition F: $F(1, 15) = 29.3$, $p \leq 0.01$, condition P: $F(1, 15) = 76.5$, $p \leq 0.01$, condition M: $F(2, 30) = 12.6$, $p \leq 0.01$, Bonferroni corrected for the three comparisons). On average, peripheral (PT) targets were fixated with a shorter latency after the appearance than distractors (D) and than foveal targets (FT).

The average fixation frequency of each item type in each experimental condition is listed in table 5.3. Fixation frequency refers here to the number of fixations on each item type in comparison to the total number of fixations on all item types. If each single item was visited with the same probability, the fixation frequency would be 0.75 for distractors and 0.25 for targets (0.25 in the conditions F and P and $2 \cdot 0.125$ in the mixed condition M, cf.

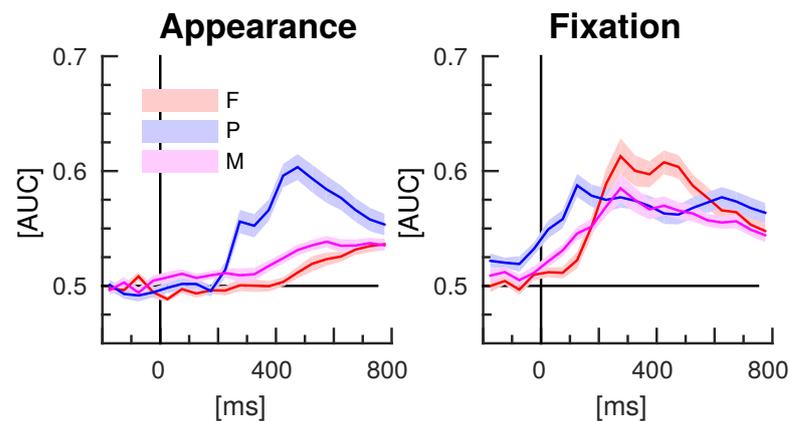


Figure 5.8: EEG classification with spatial features (separate time-intervals at all channels). Lines indicate the mean AUC-scores of the sixteen participants of the study and shaded areas stand for the standard error of the mean. [Figure from Wenzel et al., 2016c, reproduced with permission.]

Table 5.3: Fixation frequencies, averaged over participants, for foveal (FT) and peripheral targets (PT) and distractors (D) in the three experimental conditions.

	FT	PT	D
Condition F	0.289		0.711
Condition P		0.404	0.596
Condition M	0.143	0.141	0.716

section 5.2.1). Yet, the fixation frequency differed significantly from this chance level in all three conditions (two-tailed Wilcoxon signed-rank tests, $p \leq 0.01$, Bonferroni corrected for the seven comparisons). However, the effect in terms of the difference between mean value and chance level was relatively large in condition P and comparably small in condition F and M. In condition P, more peripheral targets and less distractors were fixated than what could be expected by chance, in contrast to the conditions F and M, where the fixation frequencies reflect approximately the ratio of presented foveal targets to distractors.

The duration and distance of the first saccades towards foveal (FT) vs. peripheral targets (PT) vs. distractors (D) differed significantly in the conditions F and P but not in M (cf. table 5.4). The duration and distance of the respective following saccades starting at the three item types (FT/PT/D) differed significantly in the conditions F and M but not in P. The statistics were calculated with one-way repeated measures analyses of variance and Bonferroni corrected for the three (F, P and M) tests each.

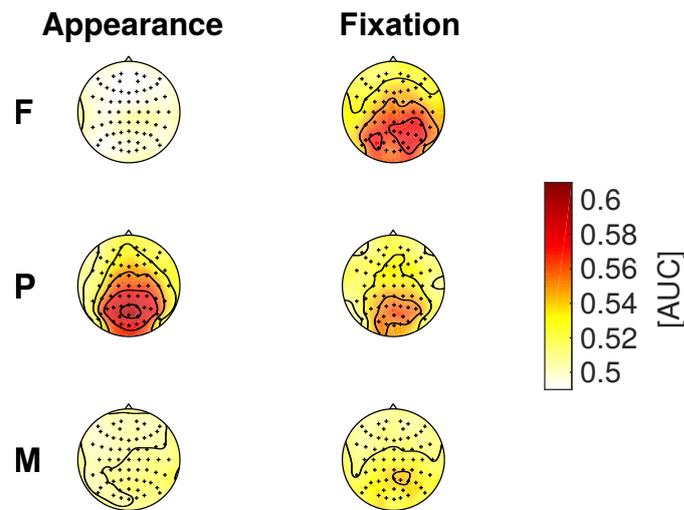


Figure 5.9: EEG classification with temporal features (entire time-series of separate channels). Average AUC-scores of the sixteen participants are presented in colour code as scalp maps. [Figure from Wenzel et al., 2016c, reproduced with permission.]

5.4 Discussion

5.4.1 Compliance check

The comparably large percentages of correct responses and the small absolute differences between response and true number of targets document that the participants were able to complete the task. The task performance was better in the experimental condition P than in condition F, where the targets were less salient and apparently missed more likely. The result of the mixed condition M, where both types of target items were presented, was situated in between the results of P and F (cf. section 5.3.1).

5.4.2 Target estimation with EEG and eye tracking features

Spatio-temporal patterns present in the neural data and features of the eye gaze were exploited to estimate which items displayed on the screen were relevant (targets) in this search task with unrestrained eye movements. Both EEG and eye tracking data contained information that made it possible to discriminate targets and distractors in all three experimental conditions (cf. table 5.1 in section 5.3.2). Crucially, the classification performance was significantly better than chance also in condition M, which modelled the more realistic scenario of a mixed item salience. Mixed salience leads to a variable timing of target recognition (cf. section 5.4.3), which is a possible reason for the lower classification performance in condition M in comparison to the conditions F and P, where target items of only one type were presented respectively. The multimodal classification of EEG and eye tracking features resulted in a better performance than when either one or the other modality was used alone (cf. section 5.3.2). Thus, the two modalities apparently contain complementary information for relevance estimation.

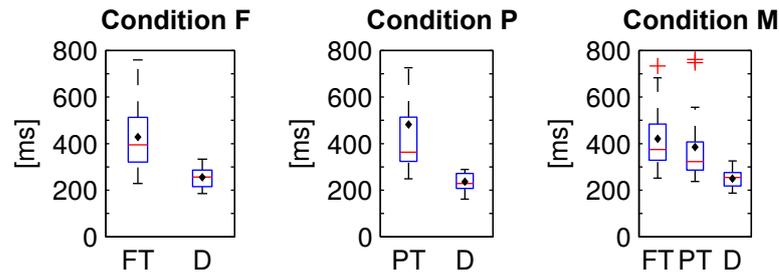


Figure 5.10: Fixation durations of foveal (FT) and peripheral targets (PT) and distractors (D) in the three experimental conditions. Average values were computed per subject and displayed as box plots. Black diamonds indicate the respective mean over participants, red lines the median, blue boxes the 25th and 75th percentiles and whiskers the range – excluding outlier participants that are marked by red plus signs. [Figure from Wenzel et al., 2016c, reproduced with permission.]

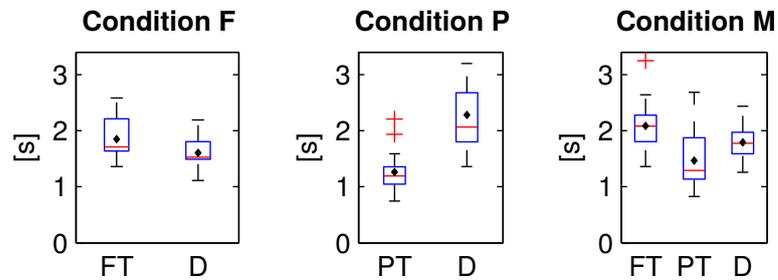


Figure 5.11: Latencies between first appearance and first fixation of foveal (FT) and peripheral targets (PT) and distractors (D) in the three experimental conditions. [Figure from Wenzel et al., 2016c, reproduced with permission.]

Eye movements are often avoided or at least constrained in EEG experiments because they can result in artefacts that deteriorate the data quality of EEG recordings [Plöchl et al., 2012] and/or constitute a confounding factor. Eye movements at a slow pace were required in recent investigations on EEG and eye tracking in search tasks [Kaunitz et al., 2014] or only long fixations were included [Brouwer et al., 2013] in order to avoid contaminations by eye movements during the interval of the late positive component. In a third study, the eye movements were otherwise constrained because the subjects had to press a key on the keyboard while fixating on the target, or to maintain the fixation on the target for at least one second [Dias et al., 2013]. Yet, restricting the eye gaze would be impractical for most real applications. For this reason, the present experimental setting was as close as possible to an application scenario. The subjects could look around without any constraints. In order to check if neural signals were indeed the basis of the previously presented EEG classification results, the classifications were additionally performed with features from the electrooculogram only. In this way, it could be tested whether the EEG results can be explained alone by the differences in the eye movements for targets and distractors, conveyed by eye artefacts to the EEG data. The EOG classification results did not exceed the chance level significantly (cf. table 5.1). Hence, neural

Table 5.4: Average duration in milliseconds (a) and distance in pixels (b) of the saccades towards foveal (FT) and peripheral targets (PT) and distractors (D). The duration (c) and distance (d) of the respective following saccades are given below. The results of the statistical comparisons between FT vs. PT vs. D are listed in the columns F , df and p (one-way repeated measures analyses of variance). Some fields of the table remain empty because FT were absent in condition P, and PT were absent in condition F.

	Condition	FT	PT	D	F	df	p
a)	F	47.5		48.2	16.3	(1,15)	≤ 0.01
	P		47.2	50.3	52.7	(1,15)	≤ 0.01
	M	47.5	47.5	47.9	1.38	(2,30)	> 0.01
b)	F	189		197	24.5	(1,15)	≤ 0.01
	P		178	208	27.4	(1,15)	≤ 0.01
	M	189	189	194	1.87	(2,30)	> 0.01
c)	F	46.6		48.3	15.3	(1,15)	≤ 0.01
	P		49.5	50.4	4.48	(1,15)	> 0.01
	M	46.3	46.5	48.1	8.04	(2,30)	≤ 0.01
d)	F	179		198	23.4	(1,15)	≤ 0.01
	P		200	210	3.71	(1,15)	> 0.01
	M	177	178	198	17.5	(2,30)	≤ 0.01

signals provided presumably the information to classify between target and distractor EEG epochs. Compare also the discussion in section 5.4.3 and 5.4.5. Classification results using EOG and eye tracking data differ presumably because the features for the EOG classification were extracted in the same way like for the EEG classification. The fixation durations were not estimated from the EOG signal.

Visual recognition could happen both in foveal and in peripheral vision in the mixed condition M. Two additional analyses of this condition were conducted (the results are listed in table 5.2):

- *Combined classifier.* Letting two classifiers learn the patterns of foveal and peripheral recognition individually, and applying them in combination, slightly improved the average performance in comparison to the standard classification (compare table 5.2, first row, with table 5.1, row ‘EEG’, column ‘M’). However, this improvement was not significant, and, therefore, it can not be stated that the combined classifier was better suited to cope with the temporal variability of neural processes related to target recognition (cf. figure 5.7, column ‘Fixation’) than the standard classifier, which did not take the variable stimulus salience into special consideration. Both parts of the combined classifier used features from fixation-aligned EEG epochs – even though appearance-aligned EEG epochs seem to be particularly suited for peripheral target detection (cf. figure 5.6 and 5.7). Yet, fixation-aligned features were almost equally suited for classification in the ex-

perimental condition P (table 5.1, row ‘EEG’, column ‘P’) as appearance-aligned features (cf. last paragraph of section 5.3.2) and fixation-aligned EEG epochs are presumably available more frequently in an application scenario, while the popping up of items in peripheral vision is rather specific for the experiment presented here.

- *Split analysis.* Either foveal or peripheral targets were classified against distractors using fixation-aligned EEG epochs from condition M only. This approach can serve as upper bound reference of what could be achievable, if it was known whether a target can be recognised in peripheral vision or not. This knowledge can not be expected in a realistic setting. For foveal targets, classification performance improved in comparison to the standard analysis (compare table 5.2, second row, with table 5.1, row ‘EEG’, column ‘M’) – probably due to the reduced temporal variability of the neural response (cf. figure 5.7, column ‘Fixation’). The result was comparable to the classification of fixation-aligned EEG epochs in condition F (cf. table 5.1, row ‘EEG’, column ‘F’). However, classifying only peripheral targets versus distractors did not result in an improvement in comparison to the standard analysis.

Appearance-aligned EEG features. Information was present in the EEG data about whether a target or a distractor item had just appeared in peripheral vision, in all experimental conditions. Classification performance was presumably better in condition P than in the conditions F and M, because in condition P peripheral detection was facilitated by the stimulus design. This type of prediction is relatively specific for the gaze contingent stimulus presentation (where items appeared in peripheral vision, cf. section 5.2.1). In contrast, the prediction based on fixation-aligned EEG epochs can be more widely applied in a human-computer interaction setting and was, therefore, main focus of the target estimation presented here. Yet, the analysis of appearance-aligned EEG epochs made it possible to check if peripheral versus foveal target recognition was experimentally induced indeed (compare also the next section 5.4.3).

Please note that AUC-scores based on the predictions of single EEG epochs can not be directly compared with the class selection accuracies which are typically reported in the literature about brain-computer interfaces. A ‘Matrix-’ or ‘Hex-O-Speller’, for instance, usually combines several sequences of several classifications for letter selection, which leads to an accumulation of evidence [cf. figure 7 and, respectively, figure 4 in Treder and Blankertz, 2010; Acqualagna and Blankertz, 2013].

5.4.3 Characteristics of target and distractor EEG epochs

Target and distractor EEG epochs were class-wise averaged and differences between the two classes were statistically assessed in order to understand the underlying reasons for the results of the classifications and to gain insight into the neural correlates of peripheral and foveal target recognition. Characteristic patterns were present in the neural data depending on whether a target or a distractor was perceived (cf. figure 5.6 and 5.7 in section 5.3.3). Their spatio-temporal dynamics suggest that the presence of the P300 component (also called P3)

differed between the two classes. This component is a positive deflection of the ERP at around 300 ms (or later) after stimulus presentation and is known to be expressed more pronounced for stimuli that are being paid attention to (here: targets) than for non-relevant stimuli (here: distractors) [Picton, 1992; Polich, 2007]. The findings of other studies with search tasks could be reproduced, where a late positive component, probably the P300, differed between fixations of targets and distractors [Brouwer et al., 2013; Kaunitz et al., 2014; Devillez et al., 2015].

The salience of target discriminative information was varied in the experiment. Accordingly, target recognition could happen either immediately after item appearance in peripheral vision, or not until the item was fixated and in foveal vision, which was reflected in the neural data as follows:

- Clear differences between appearance-aligned target vs. distractor EEG epochs were found in condition P in contrast to condition F (cf. figure 5.6 and 5.7, column ‘Appearance’) because only peripheral targets could be recognised directly after their appearance in peripheral vision. The mixed condition M was designed with the objective to model the uncertainty of a more realistic setting where recognition can happen both in foveal and in peripheral vision. Here, both types of targets were presented and, consequently, a superposition was found of the effects from condition F and P.
- Peripheral targets could be recognised already before fixation onset in contrast to foveal targets. For this reason, differences between target and distractor EEG epochs were found in condition P at earlier time points, with respect to the fixation-onset, than in condition F (cf. figure 5.6 and 5.7, column ‘Fixation’). As it can be expected, condition M represents a mixture of the effects from condition F and P.

These findings match the results of the classifications with spatial features, which had the objective to learn how the neural correlate of target recognition evolves over time after item appearance or fixation, while exploiting the multivariate nature of the multichannel EEG data (cf. figure 5.8).

Midline electrodes, mainly at central, parietal and occipital positions, were most discriminative (cf. figures 5.7 and 5.9). Hence, the results indicate that classification is not based on eye movements or facial muscle activity. These would cause higher classification performances in channels at outer positions, which are not observed here.

For figure 5.7, the EEG signals were analysed independently for all electrodes and time points. The resulting multiple testing problem was addressed with Bonferroni correction. Even though this correction is a rather conservative remedy (considering the large number of electrodes and time points), it was suited to show that the timing of the neural responses was different between conditions. The multiple testing problem could be avoided, e.g., with a general linear model with threshold free cluster enhancement [cf. Ehinger et al., 2015].

5.4.4 Eye gaze characteristics

Targets were fixated longer than distractors (cf. figure 5.10) and saccades to/from targets were quicker and shorter than those to/from distractors (cf. table 5.4). Apparently, target prediction based on eye tracking features (cf. table 5.1) was therefore possible. The longer fixation duration for targets was presumably caused by the task, because the count had to be increased by one upon the detection of a target in contrast to the recognition of a distractor that allowed the participant to directly pursue the search for the next target (cf. also the implications for the use case in the last paragraph of section 5.4.6).

The results of the eye movement analysis demonstrate that the experimental conditions effectively induced the intended effect of peripheral versus foveal target detection for the following reasons:

- Peripheral targets were fixated earlier after their first appearance than foveal targets and distractors – probably because they could be recognised as targets already in peripheral vision (cf. figure 5.11). Besides, the saccades to peripheral targets were quicker and shorter than to distractors (cf. section 5.3.4).
- The increased fixation frequency of peripheral targets in condition P (cf. table 5.3) suggests that peripheral targets could be discriminated from distractors indeed in peripheral vision. Apparently, target detection in peripheral vision resulted in saccades to targets while leaving aside distractors. In contrast, fixation frequencies almost equalled the actual percentages of targets and distractors in condition F and M. In those conditions, each item had to be fixated to determine whether it is a target (the small but significant differences between the mean fixation frequencies and the chance levels were presumably caused by the rule to early stop a repetition as soon as all targets had been fixated, cf. section 5.2.1).

5.4.5 Interference of eye movements with the EEG

The classification with EEG data (cf. section 5.4.2) was presumably successful because a late positive component, evoked by cognitive processes, differed between targets and distractors (cf. section 5.4.3). However, the hypothesis can be proposed that not cognitive processes but eye movements were responsible for the classification results. The fixation durations were shorter than the EEG epochs and shorter for targets than for distractors (cf. sections 5.4.4 and 5.2.4.2). Accordingly, the following saccade occurred still during the EEG epoch and at earlier time points in the case of targets in comparison to distractors. Saccades can interfere with the EEG because the eye is a dipole, due to activity of the eye muscles and via neural processes in the visual or motor cortex: the presaccadic spike affects the EEG signal immediately before the saccade and the lambda wave about 100 ms after the end of the saccade – both resulting in a positive deflection in particular at parietal and, respectively, parieto-occipital electrodes [Blinn, 1955; Thickbroom and Mastaglia, 1985; Thickbroom et al., 1991; Dimigen et al., 2011;

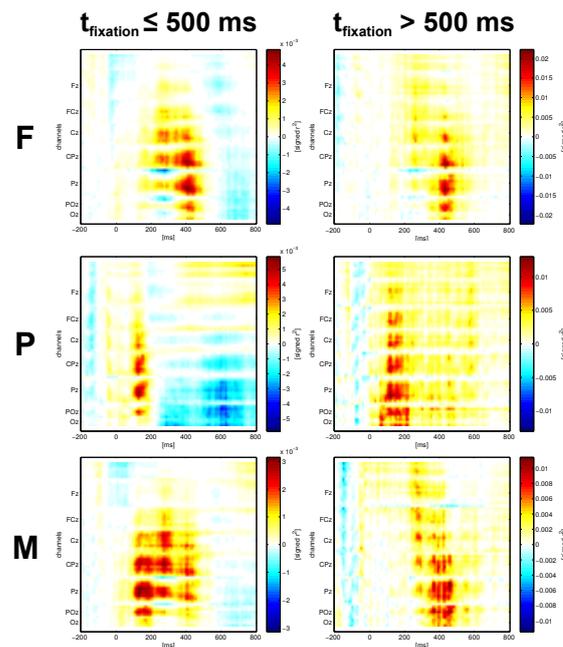


Figure 5.12: Statistical differences between target and distractor (fixation-aligned) EEG epochs with a corresponding fixation duration shorter/longer than 500 ms. The rows show the results of the experimental conditions F, P and M. Compare with figure 5.5, column ‘Fixation’, in the main text. Note that only a small proportion of the EEG epochs remained with the criterion of > 500 ms because the items were inspected quicker in the majority of the cases. [Figure from Wenzel et al., 2016c, reproduced with permission]

Plöchl et al., 2012]. This interference can not be avoided in unconstrained viewing.

Nevertheless, potentials related to cognitive processes were likely the predominant factor for the classification results and not potentials related to the following saccade for the reasons set out below. The time shift of the discriminative information (in fixation-aligned EEG epochs) between the experimental conditions F and P (cf. figure 5.7, right column) can not be explained by differences in the eye movements because the fixation durations in F and P were similar (cf. figure 5.10). A cognitive EEG component (such as the P300) is a more likely reason for the time shift because recognition was possible in condition P in peripheral vision, i.e. before fixation-onset, but only after fixation-onset in condition F.

In order to examine if the found difference patterns (cf. figure 5.7, right column) are related to a cognitive EEG component and to assure that they were not caused by the following saccade, a further test was performed and EEG epochs were selected with a corresponding fixation duration longer than 500 ms. The resulting difference patterns between target and distractor EEG epochs were similar to the case where the fixation duration was less or equal than 500 ms and to the case where all EEG epochs were used. The differences appeared again before 500 ms and thus before the following saccade (cf. figure 5.12).

Furthermore, if presaccadic spike and lambda wave were indeed responsible for the difference between target and distractor EEG epochs, a discriminative pattern could be expected, which was not observed here (cf. figure 5.7): the next saccade is expected on average 260 ms after fixation-onset for distractors and after 440 ms in the case of targets (cf. section 5.3.4 with figure 5.10). The presaccadic spikes can be assumed to occur just before these time points and the corresponding lambda waves about 150 ms later (including 50 ms for the duration of the following saccade). Both presaccadic spike and lambda wave are known to result in a parietal positivity (see above). Accordingly, the difference of target minus distractor related potentials is expected to be negative roughly at around 260 ms (distractor spike) and 410 ms (distractor λ) accompanied by a positivity at around 440 ms (target spike) and 590 ms (target λ). However, such a pattern was not observed (signed r^2 values in figure 5.7, column 'Fixation'). Instead, a P300-like pattern was predominant, which occurred earlier when target recognition in peripheral vision was possible than when foveal vision was necessary (cf. section 5.4.3).

Moreover, it was shown that EEG contains information complementary to the information from the eye tracker (cf. section 5.4.2) – even if the eye tracker measures the fixation durations very accurately in contrast to the indirect measurement with EEG. Still, EEG added information – presumably because cognitive processes were captured on top of mere effects due to the eye movements. Furthermore, the results of the individual classifications with EEG features were not correlated with the results using eye tracking features. Finally, the classification of feature vectors from the electrooculogram, which were extracted just like the EEG features, was not possible. The findings mentioned contradict the hypothesis that the difference in the fixation duration made an important contribution to the EEG classification. The long-term aim is relevance detection for tasks that are cognitively more demanding than the simple search task used here. Then, eye movements might not be sufficiently informative about the relevance any more, but accessing information about cognitive processes might be required.

5.4.6 Limitations

In view of its practical application, it has to be considered that the implicit information provided by the classifier based on EEG and eye tracking data comes with a non-negligible uncertainty. The classification performance remained considerably below an AUC of 1, which does not suffice for a reliable relevance estimation of each single item after a single fixation. This issue can be overcome by combining information derived from several uncertain predictions. Persons make several saccadic eye movements per second and, thus, EEG epochs aligned to the fixation-onsets provide a rich source of data. While the information added with each single saccade might be only a small gain, the evidence about what is relevant for a person is accumulating over time. The same strategy is followed in (ERP-based) brain-computer interfacing where typically several classifications are combined for selecting an option (e.g. a letter in a speller application).

The discriminability between targets and distractors based on EEG and eye tracking data may

to a substantial degree depend on the particular stimuli in use. Although a step towards reality was made and constraints regarding the stimuli were relaxed, there are more parameters to be considered. In this study, the salience of the target items was varied on two levels only and the presentation style was always the same (the items popped up and remained at the original position). Besides, the decision about whether an item was target or not was easy and of invariant difficulty. However, in real visual environments, a salience continuum can be expected, the presentation style can be diverse [items can move or fade in; Ušćumlić and Blankertz, 2016] and more cognitive effort can be required to evaluate the relevance of a stimulus. Thus, even more temporal variability is expected, with corresponding implications for classification. In this context, it can be noted that the temporal variability of neural responses in “real-world” environments is a problem recently addressed in the EEG literature, albeit in other respects [Meng et al., 2012; Marathe et al., 2013, 2014].

While here only effects related to the stimulus salience were examined, it is known that the task has a large influence on the visual attention [cf. Kollmorgen et al., 2010; Tatler et al., 2011]. In this experiment, target items were task relevant because they had to be counted by the subject. However, the question can be posed whether the classification algorithm learned to detect neural correlates of target recognition indeed or merely the effects of counting, which was not required for distractors (cf. chapter 6).

5.5 Conclusion

It was demonstrated that EEG and eye tracking can provide information about which items displayed on the screen are relevant in a search task with unconstrained eye movements. The specific problem addressed is that the components of the regular visual environment are typically diverse. As a consequence, the salience of target discriminative information can be variable and recognition can happen in foveal or in peripheral vision. Therefore, a variable timing – relative to the fixation-onset – of corresponding neural processes can be expected. The classification algorithm was able to cope with this uncertainty and target prediction was possible even in an experimental condition with mixed salience. Interestingly, EEG and eye tracking data were found to be complementary and neural signals related to cognitive processes were apparently captured despite of the unrestricted eye gaze. In summary, the study represents a further step for estimating the subjective relevance of items looked at, by combining methods from brain-computer interfacing and eye tracking.

5.6 Lessons learned

- The salience of the items present in the field of view has to be considered for the inference of relevance maps from EEG and eye gaze.
- BCI methods assume that the informative EEG activity is tightly time locked to the onset of *single* stimuli flashed in sequence.
- When *several* items are displayed at the same time, eye tracking can be used to relate the neural activity to the respective item looked at. The saccades (jumps of the eye gaze from one fixated position to the next) can be used as reference time points for extracting feature vectors from the continuously recorded EEG.
- The regular visual surrounding contains diverse items of a mixed salience. This salience spectrum results in a variable timing of recognition with respect to the saccades.
- The experimental results show that this temporal jitter does *not* prevent the relevance inference from EEG and eye gaze with state-of-the-art BCI methods.

6 Generalisation properties of the BCI-based relevance detector

6.1 Introduction

Methods from brain-computer interfacing can potentially infer from the electroencephalogram whether an item looked at is relevant for the individual or irrelevant. The resulting information could be used for a wide range of applications (cf. section 4.1). EEG-based relevance detection can reasonably enhance standard procedures, such as questionnaires or input with keyboard and mouse, only if the detection is performed unobtrusively in the background and exploits information that is *implicitly* contained in the signals. Inference from the EEG is not sensible if items of interest would have to be marked intentionally by a specific mental task, like silent counting in the ERP-based BCI paradigm, because this could be done as well but more reliably, economically and conveniently with a questionnaire or with the computer mouse. It is legitimate to direct the attention with a counting task towards the option of interest in a BCI writing application for paralysed people (cf. section 2.2.1 and BCI studies employing the silent counting task such as Farwell and Donchin, 1988; Sellers and Donchin, 2006; Kanoh et al., 2008; Guger et al., 2009; Brouwer and Van Erp, 2010; Treder et al., 2011; Schreuder et al., 2011; Liu et al., 2011; Manyakov et al., 2011; Acqualagna and Blankertz, 2013; An et al., 2014). However, people usually do not perform a specific mental task every time they recognise something as relevant to their respective objectives. For this reason, it was investigated if the patterns in the EEG that enable the BCI to detect relevant stimuli are closely linked to a specific task. In this case, ERP-based BCIs could be employed only for the volitional selection from different options. The alternative outcome may hold promise that the BCI can serve as a more general, implicit relevance detector.

The participants of an experimental study performed a silent counting, an arithmetic and a memory task. The tasks required the subjects to pay particular attention to certain stimuli presented on the screen. The stimulus presentation was the same in all three tasks, which allowed a direct comparison of the experimental conditions. Classifiers were trained to detect the stimuli of interest in one task, according to patterns present in the EEG signal, and were applied to data recorded when the person performed a different task. Classifiers trained with

data of one task could detect the stimuli of interest also in other tasks (irrespective of some task-related differences in the EEG). The neural activity used for the relevance prediction is apparently not strictly task specific, but likely reflects the attention allocated to the stimuli of interest.

The chapter is based on the following paper:

Wenzel, M. A., Almeida, I., and Blankertz, B. (2016b). Is neural activity detected by ERP-based brain-computer interfaces task specific? *PLoS ONE*, 11(10):1–16. doi: 10.1371/journal.pone.0165556.

6.2 Materials and methods

6.2.1 Experimental design

Thirteen people performed a silent counting, an arithmetic and a memory task. The tasks required the subject to pay particular attention to target stimuli of a colour that was randomly changed after each task repetition. The stimulus presentation was the same in all three tasks, which allowed a direct comparison of the experimental conditions. Squares in the colours magenta, yellow, red, blue and green flashed one by one for 500 ms each, interleaved by 500 ms blank screen, in a five-times-five grid in pseudo-random order and arrangement (cf. figure 6.1). The probability of the appearance of each colour was the same, such that the ratio of the random target colour to the other colours was approximately one to four, resulting in eight to thirteen targets among the 47 to 50 coloured squares in total per stimulus sequence.

Condition **C** constitutes the original version where stimuli of the target colour had to be *counted* while stimuli of other colours – the distractors – could be ignored. In the *arithmetic* task of condition **A**, ‘ten’ had to be added for targets and ‘one’ for the more frequent distractors. In condition **M**, the position of the targets on the screen had to be *memorised*. The target colour magenta appeared twice among four distractors in the short exemplary stimulus sequence given in figure 6.1. The correct result would be $1 + 1 = 2$ for condition C, $1 + 1 + 10 + 1 + 1 + 10 = 24$ for condition A and ‘row 1, column 3’ and ‘row 3, column 5’ for condition M.

The task and the random target colour were introduced before each stimulus sequence. After the presentation of the stimulus sequence, the result had to be entered with keyboard (numbers) or mouse (coordinates) and, finally, the correct answer was shown on the screen. The three tasks took turns and were repeated twenty times each (cf. figure 6.1).

Stimuli of the target colour did not stand out systematically, e.g., with respect to salience or frequency. Targets distinguished themselves only due to the preceding definition as target for the present task repetition, because each colour appeared with equal probability and the target colour was frequently changed.



Figure 6.1: Short exemplary stimulus sequence (left), experimental tasks C, A and M (centre) and sequence of the tasks and random target colours (right). The participants looked at a random stimulus sequence, where 47 to 50 squares of five colours flashed (with equal probabilities) in a grid for 500 ms each, interleaved by 500 ms blank screen. Before each stimulus sequence, the task and a random target colour were assigned. The respective target colour required a particular mental operation, depending on the task. Every participant performed task C (*counting* targets), A (*arithmetic* for targets and distractors), and M (*memorizing* target positions) twenty times each. The result had to be entered after the stimulus sequence. [Figure from Wenzel et al., 2016b, reproduced with permission.]

6.2.2 Experimental setup

Participants sat at a viewing distance of approximately eighty centimetres from the screen (refresh rate 60 Hz, resolution 1920 x 1200 pixels, size 52 cm x 32.5 cm, visual angle 33° in horizontal and 22° in vertical direction) and had access to a keyboard and a mouse. EEG signals were recorded with 64 active EEG electrodes arranged according to the international 10–20 system (*ActiCap*, *BrainAmp*, *BrainVision Recorder*, BrainProducts, Munich, Germany; sampling frequency of 1000 Hz). The ground electrode was placed on the forehead, the reference electrode on the left mastoid, one of the regular EEG electrodes on the right mastoid for later re-referencing to the linked-mastoids and another electrode below the left eye for electrooculography (EOG). Electrode impedance was set at values of 5 k Ω or less, which was possible in more than 95 % of the cases. If an optimal impedance between an electrode and the scalp could not be achieved despite considerable effort, this non-optimal impedance was accepted and the experiment was started. Maximum impedance at start time was 7 k Ω at the ground electrode, 9 k Ω at the reference electrode and 26 k Ω at a scalp electrode. Stimuli were presented with in-house software written in *Processing* (version 2.2.1, <https://processing.org>) controlled by *Matlab* (MathWorks, Natick, USA).

6.2.3 Data acquisition

Five female and eight male subjects with normal or corrected to normal vision, no report of eye or neurological diseases and ages ranging from 18 to 65 years (mean of 31.2 years) participated in the study. The tasks were introduced and trained at the beginning of the experiment of two hours. The participants gave their informed written consent to take part in the experiment. The study was approved by the ethics committee of the Department of Psychology and Ergonomics of the Technische Universität Berlin (reference BL_02_20140520).

The EEG data were re-referenced to the linked mastoids and band-pass filtered between 0.5 Hz and 40 Hz with an infinite impulse response forward-backward filter. The continuous multichannel data were segmented in one second long epochs aligned to the flashing of targets and distractors, starting at 100 ms before the respective stimulus onset. Baseline correction was applied using the data within the 100 ms long interval before stimulus onset.

The participants repeated each task twenty times and viewed 47 to 50 stimuli per task repetition. The first eight markers per repetition that indicated the stimulus onset had to be discarded due to a jitter, i.e. an imprecision, in the stimulus presentation. As result, there were 165 ± 5 target and 648 ± 13 distractor epochs (mean \pm std) available per participant and experimental condition.

6.2.4 Data analysis

6.2.4.1 Single-trial classification

The question was addressed if the neural response to target stimuli is specific to the silent counting task or if it can be also evoked by other tasks. The problem was approached by asking the subjects to perform three tasks that required to pay attention to certain stimuli. The stimuli were classified either as targets or distractors based on the immediate neural response to them. The classifiers were trained with data recorded when the subject performed one of the three tasks and tested on separate data acquired when a different task was requested. Classifiers trained in one experimental condition should be able to detect targets in different experimental conditions if the target-related neural activity is not task specific. Training and testing was performed on all possible pair-wise combinations of the three conditions. As additional reference level, every condition was inspected separately and served both for training and testing. In this case, the classification performance was assessed by splitting the data in test and training sets in a ten-fold cross-validation [Lemm et al., 2011].

Spatio-temporal features for the classifications were extracted from each EEG epoch within the interval from 100 ms to 800 ms. The EEG signal was downsampled to 20 Hz in order to improve classification performance via a reduction of the dimensionality of the features [Blankertz et al., 2011]. A 930 dimensional feature vector was obtained for each EEG epoch by concatenating the EEG potentials measured at all 62 scalp EEG channels and 15 time points within the 700 ms long epoch. Classifications were performed with regularized linear discriminant analysis where the shrinkage parameter was determined analytically [Friedman, 1989; Ledoit and Wolf, 2004; Schäfer and Strimmer, 2005]. Performance was assessed with the area under the curve (AUC) of the receiver operating characteristic [Fawcett, 2006].

Single-trial classifications were performed with all samples including trials potentially corrupted by artefacts. Accepting this challenge is expected to be useful for online operation in prospective applications. Moreover, the employed multivariate methods are able to project out artefacts of various kinds.

The previously introduced classifications were conducted separately for each participant (*within*-participants). Besides, an *across*-participants classification scheme was employed in order to investigate if a transfer of the predictor is possible between subjects, which would allow to skip a time-consuming individual calibration session (cf. section 6.4.1). For this purpose, classifiers were trained on the data of all participants but one and tested on the data of the respective withheld participant. The procedure was iterated such that the data of every participant were tested. Again, all combinations of training and testing condition were assessed. Moreover, the effect of the number of training subjects on the classification performance was determined. The data of one to twelve subjects were used to train a classifier (to discriminate between targets and distractors) that was tested on the data of each withheld participant. In this analysis, all experimental conditions were merged for the sake of conciseness and in view of the envisaged application case where the users are expected to perform various tasks. The training subjects were drawn at random if there existed several possibilities.

6.2.4.2 Spatio-temporal dynamics

Additionally, the spatio-temporal dynamics of the neural responses to the flashing of target and distractor stimuli were inspected. While the main hypothesis under investigation was tested with the classification approach detailed above, this inspection allows for a better understanding of the underlying reasons for success or failure of the classifications. The measured potentials were averaged over the single EEG epochs of all participants, separately for each experimental condition, class (targets/distractors), channel and time point.

The difference between the two classes was assessed by computing the correlation between the potentials of the single EEG epochs and the class label, 1 for targets and 0 for distractors, separately for each channel and time point. The yielded correlation coefficients were squared while retaining the original sign (signed r^2 values). Again, averages across participants were calculated. The coefficients were Fisher z-transformed before averaging to make them approximately Gaussian distributed, which was reversed after averaging to bring them back to the original unit [Silver and Dunlap, 1987]. A significance threshold was not employed in order to keep the full spatio-temporal pattern including potentially subtle differences that might be exploited by the multivariate classifier, which was introduced above.

In order to ensure a clean and undisturbed visualization of the neural responses, artefact epochs had been rejected beforehand based on a maximum-minimum criterion of $100 \mu\text{V}$ for the EEG channels and of $200 \mu\text{V}$ for the EOG channel, within the post-stimulus interval. Around 133 ± 30 target (mean \pm std) and 489 ± 150 distractor epochs remained per participant and experimental condition.

6.2.4.3 Behavioural performance

It was checked that every participant complied with the instructions and performed the tasks. For this purpose, the numbers entered and the positions clicked at were compared with the correct numbers and positions and it was statistically assessed whether the results were more accurate than it can be expected if the participants answered randomly. The distances between the correct and the entered numbers were calculated in the conditions C and A. It was assessed with Mann-Whitney U tests if the resulting distances were significantly smaller than random distances, which had been generated by shuffling the relations between correct and entered numbers a thousand times. In the condition M, the accuracies of selecting the correct target positions were computed. Mann-Whitney U tests checked if these accuracies were significantly greater than random accuracies, which had been determined by moving the targets to random positions a thousand times.

Analysis and visualization of the EEG and behavioural data were performed with Python (version 3.5.2, <http://www.python.org>), the MNE-Python software, pandas, scikit-learn and seaborn [Gramfort et al., 2013, 2014; McKinney, 2011, 2010; Pedregosa et al., 2011; Waskom et al., 2015].

6.3 Results

6.3.1 Single-trial classification

Fig 6.2 displays the results of the within-participant classifications of target versus distractor EEG epochs. The classification performance was assessed with the AUC. This metric represents both the sensitivity and the specificity of the classifier and is insensitive to class imbalances [Fawcett, 2006]. An AUC of 0.5 constitutes the chance level of the classification. For all combinations of training and testing conditions and for every participant, the AUC was consistently better than it can be expected from random guessing. Wilcoxon signed-rank tests showed that the results were on the population level significantly above an AUC of 0.5 ($p \leq 0.05$, Bonferroni corrected for the nine combinations of training and testing conditions).

The cross-validation results (values on the diagonal of the matrix in figure 6.2) might not be directly compared with the results obtained by training on one condition and testing on a different condition (on the off-diagonal of the matrix in figure 6.2).

Fig 6.3 displays the results of the classifications across-participants. Classification performance was on the population level significantly better than chance in all cases but one (C \rightarrow M, Wilcoxon statistic as above).

Data of more participants used for the classifier training resulted in a better performance when transferred to a different participant (cf. figure 6.4; the three conditions were merged for this analysis as motivated in section 6.2.4.1). The number of training subjects was significantly

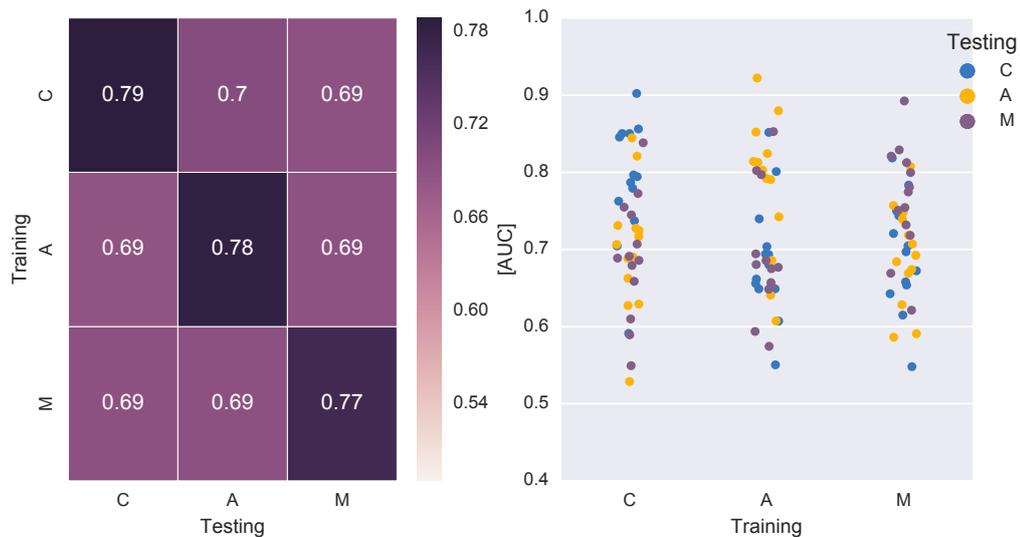


Figure 6.2: Average (left) and single participant (right) results of the classifications *within*-participants for all combinations of training and testing condition, measured as area under the curve of the receiver operating characteristic. All results were on the population level significantly better than random guessing ($p \leq 0.05$). [Figure from Wenzel et al., 2016b, reproduced with permission.]

correlated with the AUC (correlation coefficients were calculated for every subject, average: $\rho = 0.50$, t-test: $p \leq 0.05$). Nevertheless, a ceiling effect can be observed for $n \geq 6$.

6.3.2 Spatio-temporal dynamics

The spatio-temporal dynamics of the neural responses to the flashing of target and distractor stimuli are visualised in figure 6.5, 6.6 and 6.7. Averages across participants are displayed separately for the conditions C, A and M. Figure 6.5 shows the time course of the EEG potential measured at frontal, central and parietal positions along the midline of the head. Figure 6.6 depicts the time courses at all electrodes in colour code, separately for targets (top) and distractors (centre). The lower row shows the difference between the two classes. Figure 6.7 presents the data as scalp topographies.

6.3.3 Behavioural performance

Every participant entered numbers and clicked at positions that were significantly more accurate as it can be expected by chance ($p \leq 0.05$).

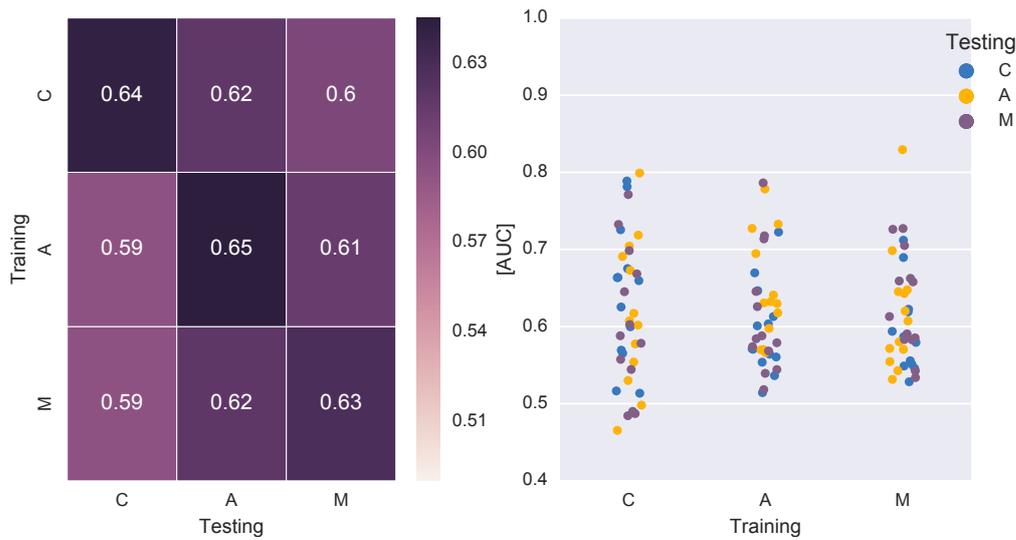


Figure 6.3: Average (left) and single participant (right) results of the classifications *across*-participants. The results were on the population level significantly better than chance, except in one case (C → M). [Figure from Wenzel et al., 2016b, reproduced with permission.]

6.4 Discussion

6.4.1 Single-trial classification

EEG epochs that were either aligned to targets or to distractors could be discriminated significantly better than it can be expected by chance (AUC of 0.5) for all combinations of training and testing conditions (within-participant classifications; cf. figure 6.2). Discrimination based on EEG data was not only possible in the classic counting variation (C) but also when both targets and distractors required arithmetic (A) or when the positions of the targets had to be memorised (M).

Each classifier could predict targets in every experimental condition and not only in the condition where it had been trained. This successful transfer suggests that a substantial part of the neural activity evoked by targets is neither specific to the silent counting, nor to the arithmetic, nor to the memory task. Both the target recognition itself, as a result of the attention allocation, and the augmented cognitive effort are equally plausible causes for the findings, because targets required a more demanding task than distractors (at least in the condition C and M where distractors could be simply ignored).

Tailoring the classifier to each individual person, as it is typically done in BCI experiments, would be a hindering factor for the application in human-computer interaction. A time-consuming calibration session constitutes a hurdle for the users to adopt EEG-based technology for the every-day interaction with a computer. Interestingly, however, it was possible to skip the individual classifier training and predict the task-relevant stimuli with a classifier that was trained on the data of other participants (across-participants classifications; cf. figure 6.3)

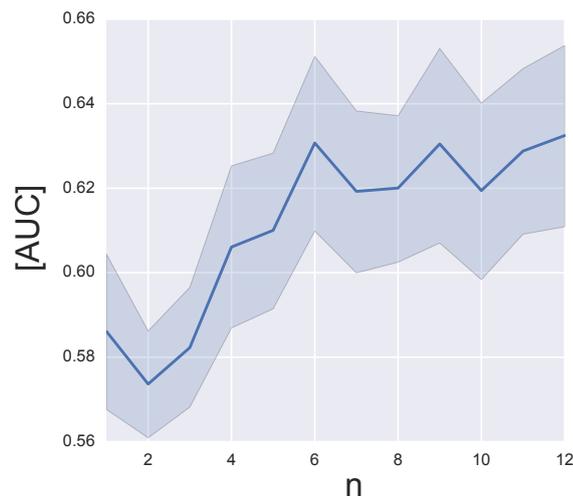


Figure 6.4: Performance (AUC) of the classification across-participants depending on the number (n) of participants used to train the classifier. The three experimental conditions were merged for this analysis (cf. section 6.2.4.1). Bootstrapping, a resampling method, was used to estimate the 68 % confidence intervals (equivalent to ± 1 standard deviation in the Gaussian case) of the mean across participants [Efron, 1979]. [Figure from Wenzel et al., 2016b, reproduced with permission.]

even if the performance was significantly inferior ($p \leq 0.05$, Wilcoxon signed-rank test) to the classification within-subjects (cf. figure 6.2). Acquiring data of more participants improved the predictive performance until a ceiling level was reached for $n \geq 6$. (cf. figure 6.4). Transfer learning methods could further improve the transferability between subjects [Lu et al., 2009; Fazli et al., 2009, 2011, 2015; Kindermans et al., 2014; Jayaram et al., 2016; Koyamada et al., 2015].

6.4.2 Spatio-temporal dynamics

The patterns in the neural data that allow for differentiating between targets and distractors were inspected in order to uncover the reason for the successful classifications. Targets evoked, in all experimental conditions, an augmented late positive component in comparison to distractors in particular at the midline centroparietal and parietal electrodes (cf. figure 6.5, 6.6, 6.7), which is typical for the P300 wave [Sutton et al., 1965; Picton, 1992; Polich, 2007].

Some differences between the conditions can be noted (cf. figures 6.5, 6.6 and 6.7): condition C, the classic variation with silent counting, featured a comparably large difference between the potentials evoked by targets and distractors. Condition A shows a comparably large late positive deflection for distractors. In this condition, all stimuli including the distractors required arithmetic and, thus, a certain amount of attention and neural processing. Finally, the discriminative neural activity lasted longer in A and M than in C (cf. figure 6.7). Presumably, the memory encoding was more variable in time in these two conditions.

6.4.3 Behavioural performance

The behavioural results show that all participants complied with the instructions and performed the tasks.

6.4.4 Limitations

Single stimuli popped up in succession in this experiment. However, it can be expected that several words or pictures are shown in parallel in a more realistic setting. The combination of EEG with an eye tracker would make it possible to relate the neural activity to each picture or word (cf. chapters 4 and 5, and Brouwer et al., 2013; Kaunitz et al., 2014; Kauppi et al., 2015; Wenzel et al., 2016c; Ušćumlić and Blankertz, 2016; Finke et al., 2016). Eye movements towards the items could be used as time points of reference for the EEG segmentation in epochs, instead of the onset of stimuli popping up on the screen. With this approach, a relevance score could be assigned to every item displayed.

It was demonstrated that the detectable neural activity evoked by targets is not specific to any of the three well-defined tasks employed in the experiment. However, it still has to be shown that relevance information can be collected implicitly in the background during the ‘natural’ interaction with a computer in the absence of precisely defined tasks.

Moreover, the stimuli used here were squares and differed only in their colour. The decision if a stimulus was a target was simple and could be performed immediately. In contrast, various pictures and words can be presented on the screen in a realistic scenario (cf. chapters 4 and 5). The decision if a picture or a word is of interest can need sometimes less and sometimes more time. Accordingly, a variable latency of the neural response can be expected, which makes relevance estimation based on neurophysiological data more difficult [cf. chapter 5 and Wenzel et al., 2016c; Ušćumlić and Blankertz, 2016].

All stimuli were similar with respect to their salience in this experiment. Yet, in a more realistic scenario, particularly salient but not necessarily relevant stimuli could elicit a passive P300, which would result in false positive estimates (even though the passive P300 is evoked rather by auditory stimuli than by visual stimuli) [Squires et al., 1975; Bennington and Polich, 1999; Jeon and Polich, 2001].

6.5 Conclusion

Based on EEG data, screen content could be classified as task relevant or irrelevant, even when different mental operations were performed than during classifier training. The results suggest that the neural activity detected by the classifiers is not strictly task specific, at least under the controlled conditions of this experimental study. This outcome may hold promise for expanding the range of application of BCI methods towards a more general detection of relevance in situations where the people do not perform a specific task each time they

recognise something which attracts their attention.

6.6 Lessons learned

- 'Relevance' has to be artificially created in experiments. For this purpose, an intrinsic interest can be mimicked with a mental task (e.g. counting all relevant items).
- Focusing the attention in this way is legitimate for a BCI application for communication, but not for the intended *implicit* relevance detection. Potentially, neural activity is detected that merely corresponds to the specific mental task, but not to the subjective experience of considering something as relevant.
- It was evaluated if the neural activity detected by the multivariate classifiers in the EEG is task specific, or – alternatively – if generalisation over different scenarios is possible.
- The BCI-based relevance detector demonstrated good generalisation properties. This result may indicate that the BCI does not rely on a specific mental task, which is promising for the intended relevance detection.

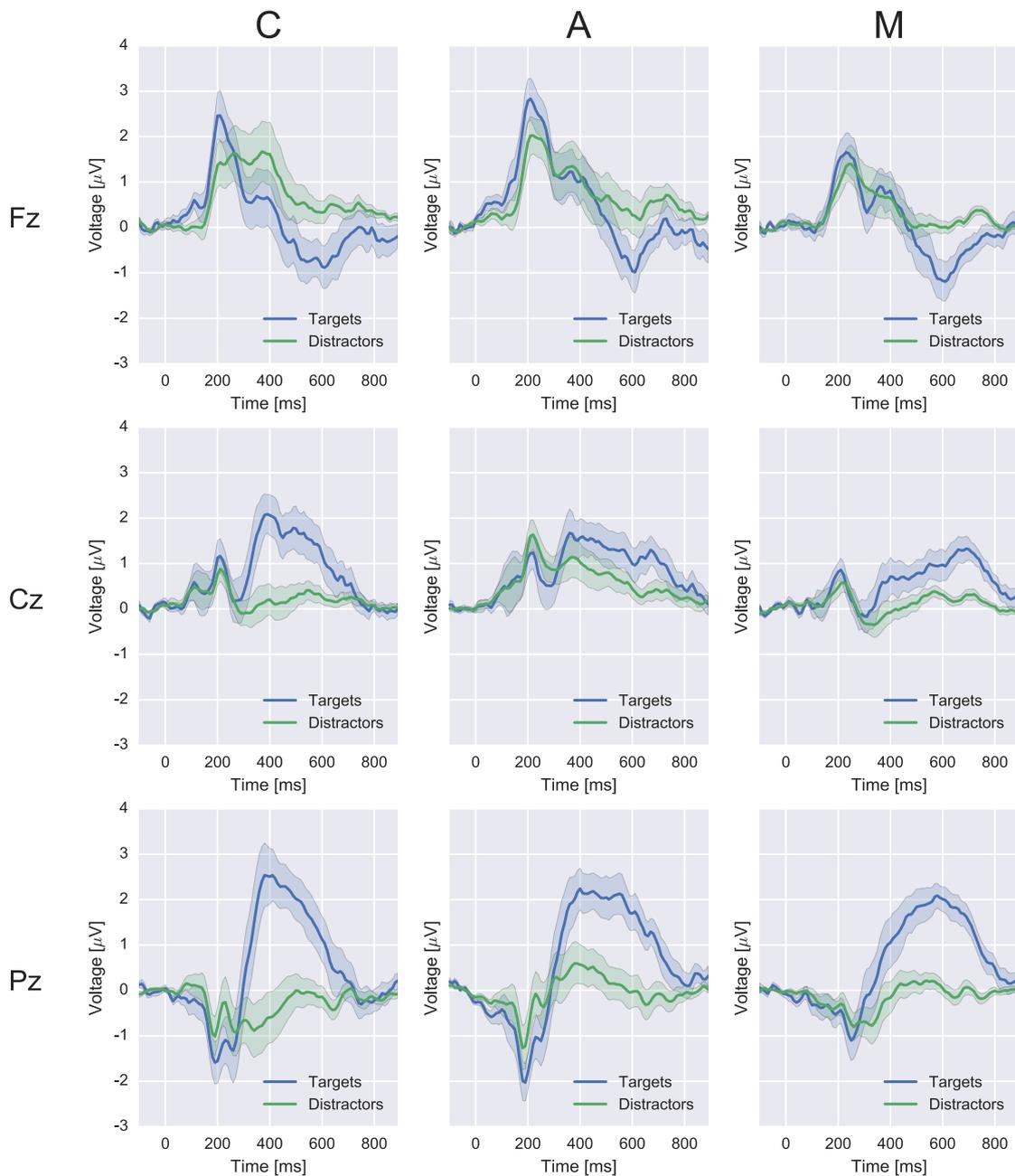


Figure 6.5: Time courses of the EEG responses to targets and distractors at the midline electrodes Fz, Cz and Pz in the experimental conditions C, A and M (averages over all epochs of all subjects). The respective stimulus-onset is situated at $t = 0$ ms. The 68 % confidence intervals were calculated with bootstrapping. [Figure from Wenzel et al., 2016b, reproduced with permission.]

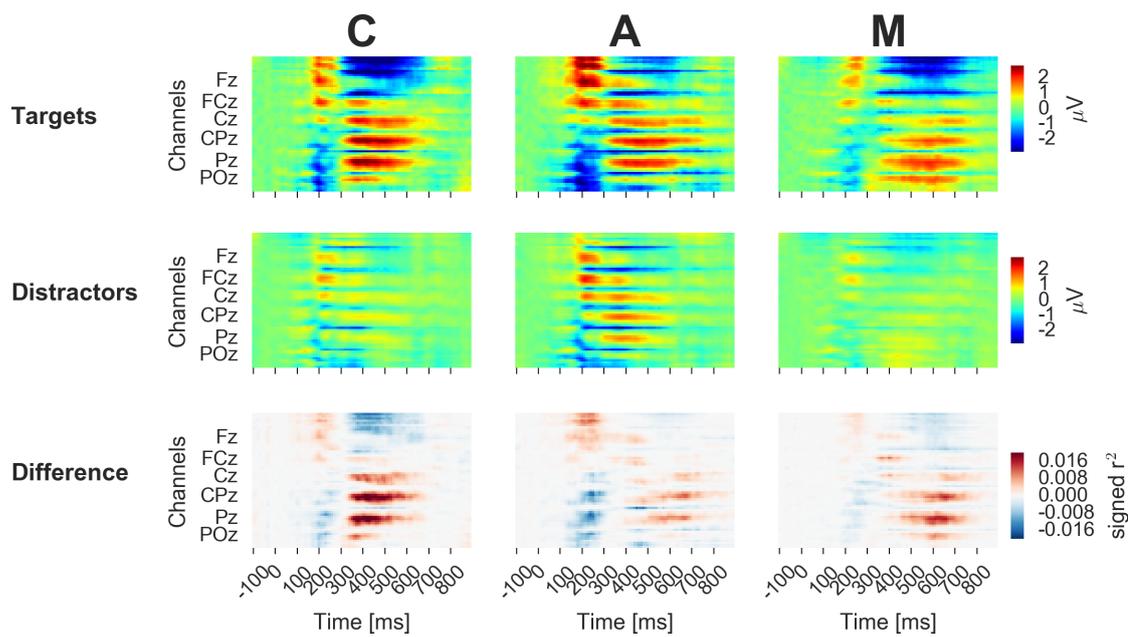


Figure 6.6: Spatio-temporal EEG activity in the three experimental conditions (mental tasks C, A and M). EEG responses to target and distractor stimuli and the corresponding differences are visualised separately (top, centre, bottom). [Figure from Wenzel et al., 2016b, reproduced with permission.]

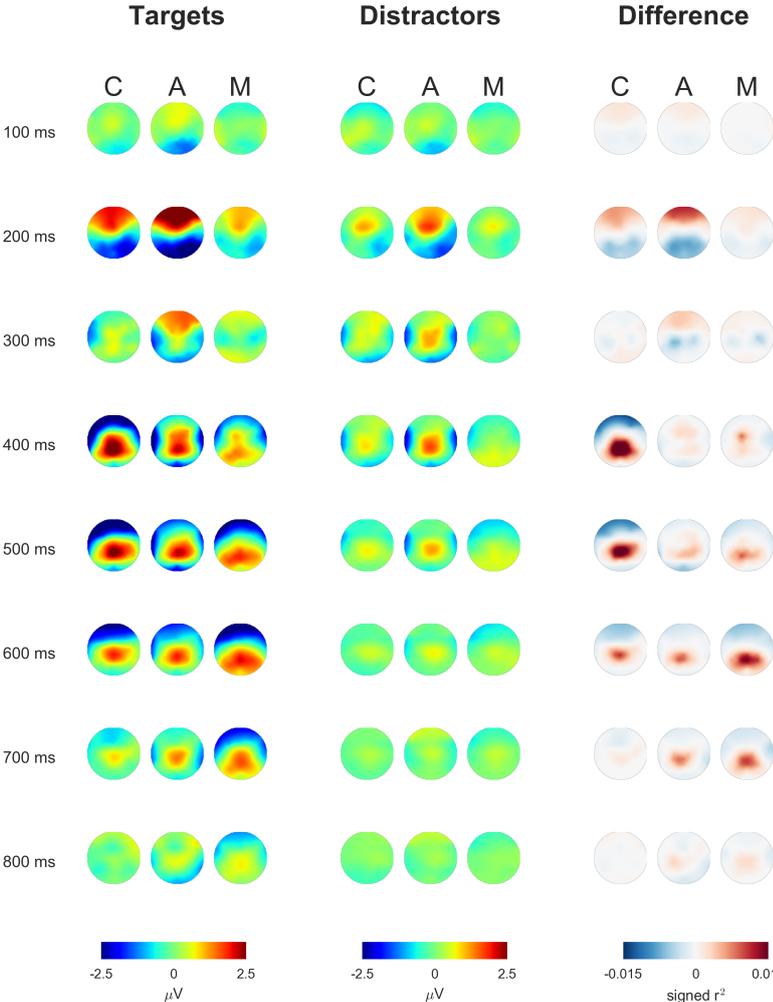


Figure 6.7: Spatio-temporal EEG activity in the three experimental conditions (mental tasks C, A and M) depicted as scalp topographies (head from above with the nose on top, average values over 50 ms long intervals around 100 ms, 200 ms, . . . , 800 ms post-stimulus-onset). Responses to target and distractor stimuli, and the corresponding differences, are visualised separately (left, centre, right). [Figure from Wenzel et al., 2016b, reproduced with permission.]

7 Discussion

7.1 Summary

Signals from the brain contain valuable user-related information. In this dissertation, it was explored how this implicit information can be accessed with BCI methods. First, it was demonstrated that BCI methods can uncover a usability flaw, which could not be noticed by test persons. The results of the study suggest a remedy that can potentially improve the ease of use of the assessed device (cf. chapter 3). Second, it was shown that the subjective relevance of the elements of the visual surrounding can be estimated based on brain activity and eye movements. The resulting relevance map of the field of view makes it possible to infer the current interest of the individual person in real-time (cf. chapter 4). A variable salience does not prevent the relevance mapping with state-of-the-art BCI methods (cf. chapter 5). The BCI-based relevance detector can generalise over different objectives (tasks) of a person. This result may indicate that the captured neural activity is related indeed to the subjective experience of considering something as relevant (cf. chapter 6).

The direct observation of the neural processes in the brain offers particular benefits in comparison to standard methods for obtaining user-related information, such as introspection, subjective reports, user questionnaires, behavioural observations, or making interaction with the device or application possible (e.g. via mouse and keyboard with computer software). First, uncovering the usability flaw and suggesting a possible remedy would have not been possible with the mentioned standard methods due to the limits of human perception (cf. chapter 3). Second, the subjective relevance mapping with EEG and eye tracking may circumvent time-consuming and distracting questions concerning the individual current interest, as well as a possible response bias. Websites, devices, vehicle cockpits, or stores can potentially be optimised with the obtained relevance maps. The implicit relevance information could be aggregated in dynamic user interest profiles. Novel types of adaptive, personalised software could take the estimated user interest into account, which would enrich the standard interaction between human and computer [cf. section 4.1 and Müller et al., 2008; Blankertz et al., 2010, 2016; Eugster et al., 2014, 2016; Kauppi et al., 2015; Finke et al., 2016; Pohlmeier

et al., 2011; Zander and Kothe, 2011; Ušćumlić et al., 2013; Jangraw et al., 2014; Blankertz et al., 2016].

Multivariate methods from machine learning and signal processing were essential for extracting information hidden in the EEG signals. As discussed in the sections 2.1 and 2.3, various processes are computed in the brain at the same time and in rapid succession, by large distributed networks of single neurons, but can be measured only indirectly on the scalp, while measurement noise and artefacts corrupt the signal quality. For an optimal coverage of the complex neural activity, EEG is recorded with several electrodes at a high temporal resolution. Information present in the resulting high dimensional data can be extracted best by multivariate methods that consider the different dimensions in combination. Independent component analysis made it possible to focus specifically on the visual cortex (cf. chapter 3). Neural activity – that the test persons themselves were unaware of – was detected in the recorded signals with pattern classification (cf. chapter 3). With regularized linear discriminant analysis, the subjective experience of relevance was inferred from high dimensional, spatio-temporal EEG feature vectors (cf. chapters 4, 5 and 6). Inspecting the informative EEG patterns was important in order to avoid ‘black box’ classification. In this way, the characteristics that the models had learned could be interpreted. It was demonstrated that neural activity from the cortex was captured indeed, and not retinal activity (cf. section 3.4.3), or interfering effects from eye movements (cf. section 5.4.5).

7.2 Limitations

Several limits have to be faced when decoding user-related information from brain signals.

First, it is worth emphasising that the presented neurophysiology-based usability assessment (cf. chapter 3) is specific for the objective of this investigation. The approach can not serve as instrument for measuring neural workload in general, but is restricted to the frequency-dependent quantification of the effect of stereoscopic shutter glasses on the brain. The evaluation of other applications or devices will require different approaches or may not be accessible for an assessment with neurotechnology at all. Importantly, it remains an open question if there is a causal relation between the detectable effect of the shutter on the brain and the prevalence of visual discomfort, which can affect some viewers of stereoscopic television. The results of the study merely suggest a possible remedy for the problem under investigation. Nevertheless, further experiments are necessary for testing whether an increased shutter frequency can – in fact – reduce the prevalence of visual discomfort. Note that the question, whether detectable cortical effects are relevant for the problem solution, applies to all neurophysiology-based investigations that turn out to be more sensitive than behavioural methods.

Second, the BCI-based relevance mapping of the field of view comes with a considerable uncertainty, which can not be neglected in practice (cf. chapters 4, 5 and 6). The single estimates are rather vague, and therefore care was taken to accumulate evidence over time (cf.

chapter 4). Fortunately, humans move the eyes several times per second, which makes rapid data collection possible. However, evidence accumulation is not possible in any use case. For this reason, it is important to select appropriate use cases in the first place, and to consider the uncertainty of the relevance estimates during the development process of an application.

The potential of the BCI-based relevance detector in realistic settings, outside of the laboratory, still has to be shown – most notably, because an intrinsic interest was mimicked with mental tasks that artificially made certain stimuli (task-)relevant. By assessing the generalisation properties of the approach, it could be excluded that mere task-specific neural activity was captured (cf. chapter 6). Nevertheless, this study can serve *only as a proxy* to the question whether the subjective experience of considering something as relevant can be decoded from the EEG indeed. Furthermore, it has to be considered that only few parameters are typically varied in experimental investigations. Under realistic conditions, the neural dynamics may be far more variable, various effects can potentially interfere, and the approach may turn out to be not specific enough. Simplifying assumptions were made, e.g. by directly contrasting relevant versus irrelevant stimuli, without a fluent transition. The presented stimuli were artificial, appeared suddenly on the screen and did not move [cf. Ušćumlić et al., 2013]. The words read and items looked at were considered as independent (cf. chapters 4, 5 and 6). However, single words in sentences are syntactically interrelated, and several sentences can build a text. Likewise, realistic visual scenes contain elements that have to be interpreted in combination. Such composite concepts can not be fully captured by evaluating each element of the field of view independently. Besides, the experiments had a comparably short duration. Over longer periods, non-stationarities can be expected, e.g. due to changes in physiology or mental strategy, which can affect the brain activity patterns. The classifier learns these patterns in a calibration phase, but does not adapt to the changes, and might be misled after some time.

At present, the BCI-based relevance detector has to be recalibrated for each single person before every usage, which should be overcome for better acceptance in practice. Likewise, current EEG systems require a lengthy preparation, and the equipment is bulky, obtrusive and expensive, at least when no compromises on the signal quality are accepted. Nevertheless, recent hardware innovations and the increasing interest in wearable sensor technology may change the game (cf. section 2.1). Most certainly, not everybody is ready to consent to personalised interest profiles inferred from brain signals. Accordingly, data privacy must be guaranteed and the expected benefits must be substantial.

7.3 Outlook

The method for uncovering an imperceptible neural workload (cf. chapter 3) could be applied to similar problems, e.g. to flickering lighting systems. In this context, the sensitivity could be compared with corresponding BCI-based approaches [Porbadnigk et al., 2011; Acqualagna et al., 2015]. Furthermore, it should be assessed if the implications of the experimental findings

can indeed reduce the prevalence of visual discomfort.

In this dissertation, the foundations were laid of a BCI-based relevance detector of the field of view (cf. chapters 4, 5 and 6). In future work, it should be evaluated to what extent the relevance detector turns out to be applicable in practice. Reality could be captured more closely with novel experimental paradigms, where an intrinsic interest in an ecologically valid visual surrounding is present indeed. EEG classification with state-of-the-art methods was successful in the studies presented here. Nevertheless, the predictive performance can potentially still be improved. Curve registration methods from functional data analysis might help for better coping with the variability of the neural dynamics [Ramsay and Silverman, 2005; Marron et al., 2015, and chapter 5]. New techniques could be developed which can adapt to confounding factors to be expected in realistic settings. The impact of misleading non-stationarities in the signals could be cushioned with techniques that find ‘stationary subspaces in multivariate time series’ [von Bünau et al., 2009]. The research-grade EEG system and eye tracker used in the laboratory could be replaced with mobile equipment such as in-ear EEG and mobile eye tracking glasses. Artefact rejection [Urigüen and Garcia-Zapirain, 2015] and robust algorithms [Samek et al., 2014] could be tailored to mobile recording situations, where movement artefacts interfere, and where measurement noise can substantially deteriorate the signal quality. Transfer learning methods could render possible subject independent classifiers that circumvent the time-consuming individual calibration sessions [Lu et al., 2009; Fazli et al., 2009, 2011, 2015; Kindermans et al., 2014; Jayaram et al., 2016]. Finally, information distributed over space and time (e.g. when reading words in sentences) could be integrated in dynamic models that make it possible to infer user-related information on a higher level.

Bibliography

- Abeles, M. (1991). *Corticonics: Neural circuits of the cerebral cortex*. Cambridge University Press, Cambridge, UK.
- Acharya, U. R., Sree, S. V., Swapna, G., Martis, R. J., and Suri, J. S. (2013). Automated EEG analysis of epilepsy: A review. *Knowledge-Based Systems*, 45:147–165. doi: 10.1016/j.knosys.2013.02.014.
- Acqualagna, L. and Blankertz, B. (2013). Gaze-independent BCI-spelling using rapid serial visual presentation (RSVP). *Clinical Neurophysiology*, 124(5):901–908. doi: 10.1016/j.clinph.2012.12.050.
- Acqualagna, L., Bosse, S., Porbadnigk, A. K., Curio, G., Müller, K.-R., Wiegand, T., and Blankertz, B. (2015). EEG-based classification of video quality perception using steady state visual evoked potentials (SSVEPs). *Journal of Neural Engineering*, 12(2):026012. doi: 10.1088/1741-2560/12/2/026012.
- Acqualagna, L., Treder, M. S., and Blankertz, B. (2013). Chroma speller: Isotropic visual stimuli for truly gaze-independent spelling. In *2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)*, pages 1041–1044. IEEE. doi: 10.1109/NER.2013.6696115.
- Adrian, E. D. and Matthews, B. H. C. (1934). The Berger rhythm: Potential changes from the occipital lobes in man. *Brain*, 57:355–385.
- An, X., Höhne, J., Ming, D., and Blankertz, B. (2014). Exploring combinations of auditory and visual stimuli for gaze-independent brain-computer interfaces. *PLoS ONE*, 9(10):e111070. doi: 10.1371/journal.pone.0111070.
- Antons, J.-N., Blankertz, B., Curio, G., Möller, S., Porbadnigk, A. K., and Schleicher, R. (2010). Subjective listening tests and neural correlates of speech degradation in case of signal-correlated noise. In *Audio Engineering Society Convention 129*, pages 8206:1–4, San Francisco, USA. Audio Engineering Society.
- Antons, J.-N., Schleicher, R., Arndt, S., Möller, S., Porbadnigk, A. K., and Curio, G. (2012). Analyzing speech quality perception using electroencephalography. *IEEE Journal of Selected Topics in Signal Processing*, 6(6):721–731. doi: 10.1109/JSTSP.2012.2191936.

Bibliography

- Aricò, P., Borghini, G., Di Flumeri, G., Colosimo, A., Bonelli, S., Golfetti, A., Pozzi, S., Imbert, J.-P., Granger, G., Benhacene, R., et al. (2016). Adaptive automation triggered by EEG-based mental workload index: A passive brain-computer interface application in realistic air traffic control environment. *Frontiers in Human Neuroscience*, 10. doi: 10.3389/fnhum.2016.00539.
- Arndt, S., Antons, J.-N., Schleicher, R., Möller, S., and Curio, G. (2012). Perception of low-quality videos analyzed by means of electroencephalography. In *Fourth International Workshop on Quality of Multimedia Experience*, pages 284–289. IEEE. doi: 10.1109/QoMEX.2012.6263836.
- Arndt, S., Wenzel, M. A., Antons, J.-N., Köster, E., Möller, S., and Curio, G. (2014). A next step towards measuring perceived quality of speech through physiology. In *Proceedings of INTERSPEECH 2014*, pages 1998–2001, Singapore.
- Ayaz, H., Onaral, B., Izzetoglu, K., Shewokis, P. A., McKendrick, R., and Parasuraman, R. (2013). Continuous monitoring of brain dynamics with functional near infrared spectroscopy as a tool for neuroergonomic research: Empirical examples and a technological development. *Frontiers in Human Neuroscience*, 7:871. doi: 10.3389/fnhum.2013.00871.
- Ayaz, H., Shewokis, P. A., Bunce, S., Izzetoglu, K., Willems, B., and Onaral, B. (2012). Optical brain monitoring for operator training and mental workload assessment. *NeuroImage*, 59(1):36–47. doi: 10.1016/j.neuroimage.2011.06.023.
- Azevedo, F. A. C., Carvalho, L. R. B., Grinberg, L. T., Farfel, J. M., Ferretti, R. E. L., Leite, R. E. P., Filho, W. J., Lent, R., and Herculano-Houzel, S. (2009). Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *Journal of Comparative Neurology*, 513(5):532–541. doi: 10.1002/cne.21974.
- Baccino, T. and Manunta, Y. (2005). Eye-fixation-related potentials: Insight into parafoveal processing. *Journal of Psychophysiology*, 19(3):204–215. doi: 10.1027/0269-8803.19.3.204.
- Baker, C. H. (1970). A study of the Sherrington effect. *Perception & Psychophysics*, 8(6):406–410. doi: 10.3758/BF03207034.
- Baldwin, C. L. and Penaranda, B. N. (2012). Adaptive training using an artificial neural network and EEG metrics for within- and cross-task workload classification. *NeuroImage*, 59(1):48–56. doi: 10.1016/j.neuroimage.2011.07.047.
- Bando, T., Iijima, A., and Yano, S. (2012). Visual fatigue caused by stereoscopic images and the search for the requirement to prevent them: A review. *Displays*, 33(2):76–83. doi: 10.1016/j.displa.2011.09.001.
- Beck, A. (1888). O pobudliwości różnych miejsc tego samego nerwu. In *Rozpr. Wydz. mat.-przyr. Polsk. Akad. Um*, volume 15, pages 165–195.
- Beck, A. (1890). Die Bestimmung der Localisation der Gehirn und Rückenmarkfunctionen mittelst der elektrischen Erscheinungen. *Zentralblatt für Physiologie*, 4:473–476.

- Belitski, A., Farquhar, J., and Desain, P. (2011). P300 audio-visual speller. *Journal of Neural Engineering*, 8(2):025022. doi: 10.1088/1741-2560/8/2/025022.
- Bennington, J. Y. and Polich, J. (1999). Comparison of P300 from passive and active tasks for auditory and visual stimuli. *International Journal of Psychophysiology*, 34(2):171–177. doi: 10.1016/S0167-8760(99)00070-7.
- Berger, H. (1929). Über das Elektrenkephalogramm des Menschen. *European Archives of Psychiatry and Clinical Neuroscience*, 87(1):527–570.
- Berka, C., Levendowski, D. J., Lumicao, M. N., Yau, A., Davis, G., Zivkovic, V. T., Olmstead, R. E., Tremoulet, P. D., and Craven, P. L. (2007). EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviation, Space, and Environmental Medicine*, 78(Supplement 1):B231–B244.
- Berman, S. M., Greenhouse, D. S., Bailey, I. L., Clear, R. D., and Raasch, T. W. (1991). Human electroretinogram responses to video displays, fluorescent lighting, and other high frequency sources. *Optometry & Vision Science*, 68(8):645–662.
- Bin, G., Gao, X., Wang, Y., Li, Y., Hong, B., and Gao, S. (2011). A high-speed BCI based on code modulation VEP. *Journal of Neural Engineering*, 8(2):025015. doi: 10.1088/1741-2560/8/2/025015.
- Birbaumer, N., Ghanayim, N., Hinterberger, T., Iversen, I., Kotchoubey, B., Kubler, A., Perelmouter, J., Taub, E., and Flor, H. (1999). A spelling device for the paralysed. *Nature*, 398(6725):297–298. doi: 10.1038/18581.
- Bishop, C. (2007). *Pattern Recognition and Machine Learning*. Springer, New York, USA.
- Blankertz, B., Acqualagna, L., Dähne, S., Haufe, S., Schultze-Kraft, M., Sturm, I., Ušćumlić, M., Wenzel, M. A., Curio, G., and Müller, K.-R. (2016). The Berlin Brain-Computer Interface: Progress beyond communication and control. *Frontiers in Neuroscience*, 10:530. doi: 10.3389/fnins.2016.00530.
- Blankertz, B., Curio, G., and Müller, K.-R. (2002). Classifying single trial EEG: Towards brain computer interfacing. In Dietterich, T. G., Becker, S., and Ghahramani, Z., editors, *Advances in Neural Information Processing Systems 14*, pages 157–164. MIT Press. <http://papers.nips.cc/paper/2030-classifying-single-trial-eeg-towards-brain-computer-interfacing.pdf>.
- Blankertz, B., Dornhege, G., Krauledat, M., Müller, K.-R., and Curio, G. (2007). The non-invasive Berlin Brain–Computer Interface: Fast acquisition of effective performance in untrained subjects. *NeuroImage*, 37(2):539–550. doi: 10.1016/j.neuroimage.2007.01.051.
- Blankertz, B., Lemm, S., Treder, M., Haufe, S., and Müller, K.-R. (2011). Single-trial analysis and classification of ERP components – A tutorial. *NeuroImage*, 56(2):814–825. doi: 10.1016/j.neuroimage.2010.06.048.

Bibliography

- Blankertz, B., Tangermann, M., Vidaurre, C., Fazli, S., Sannelli, C., Haufe, S., Maeder, C., Ramsey, L., Sturm, I., Curio, G., and Müller, K.-R. (2010). The Berlin Brain–Computer Interface: Non-medical uses of BCI technology. *Frontiers in Neuroscience*, 4. doi: 10.3389/fnins.2010.00198.
- Blankertz, B., Tomioka, R., Lemm, S., Kawanabe, M., and Müller, K.-R. (2008). Optimizing spatial filters for robust EEG single-trial analysis. *IEEE Signal Processing Magazine*, 25(1):41–56. doi: 10.1109/MSP.2008.4408441.
- Blinn, K. A. (1955). Focal anterior temporal spikes from external rectus muscle. *Electroencephalography and Clinical Neurophysiology*, 7:299–302.
- Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., and Babiloni, F. (2014). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neuroscience & Biobehavioral Reviews*, 44:58–75. doi: 10.1016/j.neubiorev.2012.10.003.
- Brazier, M. A. B. (1959). The EEG in epilepsy a historical note. *Epilepsia*, 1(1-5):328–336.
- Brazier, M. A. B. (1961). *A history of the electrical activity of the brain: The first half-century*. Macmillan, Oxford, UK.
- Brouwer, A.-M., Hogervorst, M. A., Van Erp, J. B. F., Heffelaar, T., Zimmerman, P. H., and Oostenveld, R. (2012). Estimating workload using EEG spectral power and ERPs in the n-back task. *Journal of Neural Engineering*, 9(4):045008. doi: 10.1088/1741-2560/9/4/045008.
- Brouwer, A.-M., Reuderink, B., Vincent, J., van Gerven, M. A. J., and van Erp, J. B. F. (2013). Distinguishing between target and nontarget fixations in a visual search task using fixation-related potentials. *Journal of Vision*, 13(3):17. doi: 10.1167/13.3.17.
- Brouwer, A.-M. and Van Erp, J. (2010). A tactile P300 brain-computer interface. *Frontiers in Neuroscience*, 4:19. doi: 10.3389/fnins.2010.00019.
- Brouwer, A.-M., Zander, T. O., van Erp, J. B. F., Korteling, J. E., and Bronkhorst, A. W. (2015). Using neurophysiological signals that reflect cognitive or affective state: Six recommendations to avoid common pitfalls. *Frontiers in Neuroscience*, 9:136. doi: 10.3389/fnins.2015.00136.
- Brundrett, G. W. (1974). Human sensitivity to flicker. *Lighting Research and Technology*, 6(3):127–143. doi: 10.1177/096032717400600302.
- Buzsáki, G., Anastassiou, C. A., and Koch, C. (2012). The origin of extracellular fields and currents—EEG, ECoG, LFP and spikes. *Nature Reviews Neuroscience*, 13(6):407–420. doi: 10.1038/nrn3241.
- Campbell, I. G. (2009). EEG recording and analysis for sleep research. *Current Protocols in Neuroscience*. doi: 10.1002/0471142301.ns1002s49.

- Carmena, J. M., Lebedev, M. A., Crist, R. E., O'Doherty, J. E., Santucci, D. M., Dimitrov, D. F., Patil, P. G., Henriquez, C. S., and Nicolelis, M. A. L. (2003). Learning to control a brain-machine interface for reaching and grasping by primates. *PLoS Biology*, 1(2). doi: 10.1371/journal.pbio.0000042.
- Caton, R. (1875). Electrical currents of the brain. *Journal of Nervous and Mental Disease*, 2(4):610.
- Chen, C., Li, K., Wu, Q., Wang, H., Qian, Z., and Sudlow, G. (2013). EEG-based detection and evaluation of fatigue caused by watching 3DTV. *Displays*, 34(2):81–88. doi: 10.1016/j.displa.2013.01.002.
- Cheng, M., Gao, X., Gao, S., and Xu, D. (2002). Design and implementation of a brain-computer interface with high transfer rates. *IEEE Transactions on Biomedical Engineering*, 49(10):1181–1186. doi: 10.1109/TBME.2002.803536.
- Cho, S.-H. and Kang, H.-B. (2012). The measurement of eyestrain caused from diverse binocular disparities, viewing time and display sizes in watching stereoscopic 3D content. In *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 23–28. doi: 10.1109/CVPRW.2012.6238904.
- Cohen, D. (1972). Magnetoencephalography: Detection of the brain's electrical activity with a superconducting magnetometer. *Science*, 175(4022):664–666.
- Cohen, D. et al. (1968). Magnetoencephalography: Evidence of magnetic fields produced by alpha-rhythm currents. *Science*, 161(3843):784–786.
- Collinger, J. L., Wodlinger, B., Downey, J. E., Wang, W., Tyler-Kabara, E. C., Weber, D. J., McMorland, A. J. C., Velliste, M., Boninger, M. L., and Schwartz, A. B. (2013). High-performance neuroprosthetic control by an individual with tetraplegia. *The Lancet*, 381(9866):557–564. doi: 10.1016/S0140-6736(12)61816-9.
- Creusere, C. D., Kroger, J., Siddenki, S. R., Davis, P., and Hardin, J. (2012). Assessment of subjective audio quality from EEG brain responses using time-space-frequency analysis. In *2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*, pages 2704–2708. IEEE.
- Dähne, S., Bießmann, F., Samek, W., Haufe, S., Goltz, D., Gundlach, C., Villringer, A., Fazli, S., and Müller, K.-R. (2015). Multivariate machine learning methods for fusing multimodal functional neuroimaging data. *Proceedings of the IEEE*, 103(9):1507–1530. doi: 10.1109/JPROC.2015.2425807.
- Dähne, S., Meinecke, F. C., Haufe, S., Höhne, J., Tangermann, M., Müller, K.-R., and Nikulin, V. V. (2014a). SPoC: A novel framework for relating the amplitude of neuronal oscillations to behaviorally relevant parameters. *NeuroImage*, 86:111–122. doi: 10.1016/j.neuroimage.2013.07.079.

Bibliography

- Dähne, S., Nikulin, V. V., Ramírez, D., Schreier, P. J., Müller, K.-R., and Haufe, S. (2014b). Finding brain oscillations with power dependencies in neuroimaging data. *NeuroImage*, 96:334–348. doi: 10.1016/j.neuroimage.2014.03.075.
- Darvas, F., Pantazis, D., Kucukaltun-Yildirim, E., and Leahy, R. (2004). Mapping human brain function with MEG and EEG: Methods and validation. *NeuroImage*, 23:S289–S299. doi: 10.1016/j.neuroimage.2004.07.014.
- De Lange Dzn, H. (1958). Research into the dynamic nature of the human fovea → Cortex systems with intermittent and modulated light. I. Attenuation characteristics with white and colored light. *Journal of the Optical Society of America*, 48(11):777. doi: 10.1364/JOSA.48.000777.
- De Vos, M., Gandras, K., and Debener, S. (2014). Towards a truly mobile auditory brain-computer interface: Exploring the P300 to take away. *International Journal of Psychophysiology*, 91(1):46–53. doi: 10.1016/j.ijpsycho.2013.08.010.
- Debener, S., Emkes, R., De Vos, M., and Bleichner, M. (2015). Unobtrusive ambulatory EEG using a smartphone and flexible printed electrodes around the ear. *Scientific reports*, 5. doi: 10.1038/srep16743.
- Devillez, H., Guyader, N., and Guérin-Dugué, A. (2015). An eye fixation-related potentials analysis of the P300 potential for fixations onto a target object when exploring natural scenes. *Journal of Vision*, 15(13):20. doi: 10.1167/15.13.20.
- Dias, J. C., Sajda, P., Dmochowski, J. P., and Parra, L. C. (2013). EEG precursors of detected and missed targets during free-viewing search. *Journal of Vision*, 13(13):13. doi: 10.1167/13.13.13.
- Dimigen, O., Kliegl, R., and Sommer, W. (2012). Trans-saccadic parafoveal preview benefits in fluent reading: A study with fixation-related brain potentials. *NeuroImage*, 62(1):381–393. doi: 10.1016/j.neuroimage.2012.04.006.
- Dimigen, O., Sommer, W., Hohlfeld, A., Jacobs, A. M., and Kliegl, R. (2011). Coregistration of eye movements and EEG in natural reading: Analyses and review. *Journal of Experimental Psychology: General*, 140(4):552. doi: 10.1037/a0023885.
- Donchin, E. (1966). A multivariate approach to the analysis of average evoked potentials. *IEEE Transactions on Biomedical Engineering*, BME-13(3):131–139. doi: 10.1109/TBME.1966.4502423.
- Dornhege, G. R., del R. Millán, J., Hinterberger, T., McFarland, D., and Müller, K.-R., editors (2007). *Toward Brain-Computer Interfacing*. MIT Press, Cambridge, MA, USA.
- Duda, R. O., Hart, P. E., and Stork, D. G. (2012). *Pattern Classification*. John Wiley & Sons, New York, USA.

- Duncan, C. C., Barry, R. J., Connolly, J. F., Fischer, C., Michie, P. T., Näätänen, R., Polich, J., Reinvang, I., and Van Petten, C. (2009). Event-related potentials in clinical research: Guidelines for eliciting, recording, and quantifying mismatch negativity, P300, and N400. *Clinical Neurophysiology*, 120(11):1883–1908. doi: 10.1016/j.clinph.2009.07.045.
- Efron, B. (1979). *Bootstrap methods: Another look at the jackknife*, volume 7. The Institute of Mathematical Statistics. doi: 10.1214/aos/1176344552.
- Ehinger, B. V., König, P., and Ossandón, J. P. (2015). Predictions of visual content across eye movements and their modulation by inferred information. *Journal of Neuroscience*, 35(19):7403–7413. doi: 10.1523/JNEUROSCI.5114-14.2015.
- Einevoll, G. T., Kayser, C., Logothetis, N. K., and Panzeri, S. (2013). Modelling and analysis of local field potentials for studying the function of cortical circuits. *Nature Reviews Neuroscience*, 14(11):770–785. doi: 10.1038/nrn3599.
- Elbert, T., Rockstroh, B., Lutzenberger, W., and Birbaumer, N. (1980). Biofeedback of slow cortical potentials. I. *Electroencephalography and Clinical Neurophysiology*, 48(3):293–301. doi: 10.1016/0013-4694(80)90265-5.
- Eugster, M. J. A., Ruotsalo, T., Spapé, M. M., Kosunen, I., Barral, O., Ravaja, N., Jacucci, G., and Kaski, S. (2014). Predicting term-relevance from brain signals. In *Proceedings of the 37th international ACM SIGIR Conference on Research & Development in Information Retrieval*, pages 425–434. Association for Computing Machinery. doi: 10.1145/2600428.2609594.
- Eugster, M. J. A., Ruotsalo, T., Spapé, M. M. A., Barral, O., Ravaja, N., Jacucci, G., and Kaski, S. (2016). Natural brain-information interfaces: Recommending information by relevance inferred from human brain signals. *arXiv preprint*. <http://arxiv.org/abs/1607.03502>.
- Farquhar, J., Blankespoor, J., Vlek, R., and Desain, P. (2008). Towards a noise-tagging auditory BCI-paradigm. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.524.7416&rep=rep1&type=pdf>.
- Farwell, L. A. and Donchin, E. (1988). Talking off the top of your head: Toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and Clinical Neurophysiology*, 70(6):510–523. doi: 10.1016/0013-4694(88)90149-6.
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8):861–874. doi: 10.1016/j.patrec.2005.10.010.
- Fazli, S., Dähne, S., Samek, W., Bießmann, F., and Müller, K.-R. (2015). Learning from more than one data source: Data fusion techniques for sensorimotor rhythm-based brain–computer interfaces. *Proceedings of the IEEE*, 103(6):891–906. doi: 10.1109/JPROC.2015.2413993.
- Fazli, S., Danóczy, M., Schelldorfer, J., and Müller, K.-R. (2011). ℓ_1 -penalized linear mixed-effects models for high dimensional data with application to BCI. *NeuroImage*, 56(4):2100–2108. doi: 10.1016/j.neuroimage.2011.03.061.

Bibliography

- Fazli, S., Mehnert, J., Steinbrink, J., Curio, G., Villringer, A., Müller, K.-R., and Blankertz, B. (2012). Enhanced performance by a hybrid NIRS–EEG brain computer interface. *NeuroImage*, 59(1):519–529. doi: 10.1016/j.neuroimage.2011.07.084.
- Fazli, S., Popescu, F., Danóczy, M., Blankertz, B., Müller, K.-R., and Grozea, C. (2009). Subject-independent mental state classification in single trials. *Neural Networks*, 22(9):1305–1312. doi: 10.1016/j.neunet.2009.06.003.
- Ferrari, M. and Quaresima, V. (2012). A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application. *NeuroImage*, 63(2):921–935. doi: 10.1016/j.neuroimage.2012.03.049.
- Finke, A., Essig, K., Marchioro, G., and Ritter, H. (2016). Toward FRP-based brain-machine interfaces – Single-trial classification of fixation-related potentials. *PLoS ONE*, 11(1):e0146848. doi: 10.1371/journal.pone.0146848.
- Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7(2):179–188. doi: 10.1111/j.1469-1809.1936.tb02137.x.
- Frey, J., Appriou, A., Lotte, F., and Hachet, M. (2016). Classifying EEG signals during stereoscopic visualization to estimate visual comfort. *Computational Intelligence and Neuroscience*, 2016:7. doi: 10.1155/2016/2758103.
- Frey, J., Pommereau, L., Lotte, F., and Hachet, M. (2014). Assessing the zone of comfort in stereoscopic displays using EEG. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '14, pages 2041–2046, New York, USA. Association for Computing Machinery. doi: 10.1145/2559206.2581191.
- Friedman, J. H. (1989). Regularized discriminant analysis. *Journal of the American Statistical Association*, 84(405):165. doi: 10.2307/2289860.
- Fukunaga, K. (1972). *Introduction to statistical pattern recognition*. Academic Press, San Diego, CA, USA.
- Gaebler, M., Biessmann, F., Lamke, J.-P., Müller, K.-R., Walter, H., and Hetzer, S. (2014). Stereoscopic depth increases intersubject correlations of brain networks. *NeuroImage*, 100:427–434. doi: 10.1016/j.neuroimage.2014.06.008.
- Geuze, J., Farquhar, J., and Desain, P. (2014). Towards a communication brain computer interface based on semantic relations. *PLoS ONE*, 9(2):e87511. doi: 10.1371/journal.pone.0087511.
- Geuze, J., van Gerven, M. A. J., Farquhar, J., and Desain, P. (2013). Detecting semantic priming at the single-trial level. *PLoS ONE*, 8(4):e60377. doi: 10.1371/journal.pone.0060377.
- Gevens, A. and Smith, M. E. (2003). Neurophysiological measures of cognitive workload during human-computer interaction. *Theoretical Issues in Ergonomics Science*, 4(1-2):113–131. doi: 10.1080/14639220210159717.

- Golenia, J.-E., Wenzel, M. A., and Blankertz, B. (2015). Live demonstrator of EEG and eye-tracking input for disambiguation of image search results. In Blankertz, B., Jacucci, G., Gamberini, L., Spagnoli, A., and Freeman, J., editors, *Symbiotic Interaction*, volume 9359 of *Lecture Notes in Computer Science*, pages 81–86. Springer International Publishing. doi: 10.1007/978-3-319-24917-9_8.
- Golenia*, J.-E., Wenzel*, M. A., Bogojeski, M., and Blankertz, B. (2017). Implicit relevance feedback from electroencephalography and eye tracking in image search. *Submitted*. *) Equal contribution.
- Goverdovsky, V., Looney, D., Kidmose, P., and Mandic, D. P. (2016a). In-ear EEG from viscoelastic generic earpieces: Robust and unobtrusive 24/7 monitoring. *IEEE Sensors Journal*, 16(1):271–277. doi: 10.1109/JSEN.2015.2471183.
- Goverdovsky, V., von Rosenberg, W., Nakamura, T., Looney, D., Sharp, D. J., Papavassiliou, C., Morrell, M. J., and Mandic, D. P. (2016b). Hearables: Multimodal physiological in-ear sensing. *arXiv preprint arXiv:1609.03330*. <https://arxiv.org/abs/1609.03330>.
- Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., and Hämäläinen, M. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, 7(267). doi: 10.3389/fnins.2013.00267.
- Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Parkkonen, L., and Hämäläinen, M. S. (2014). MNE software for processing MEG and EEG data. *NeuroImage*, 86:446–460. doi: 10.1016/j.neuroimage.2013.10.027.
- Grech, R., Cassar, T., Muscat, J., Camilleri, K. P., Fabri, S. G., Zervakis, M., Xanthopoulos, P., Sakkalis, V., and Vanrumste, B. (2008). Review on solving the inverse problem in EEG source analysis. *Journal of NeuroEngineering and Rehabilitation*, 5(1):1–33. doi: 10.1186/1743-0003-5-25.
- Grinvald, A., Frostig, R. D., Lieke, E., and Hildesheim, R. (1988). Optical imaging of neuronal activity. *Physiological Reviews*, 68(4):1285–1366.
- Grozea, C., Voinescu, C. D., and Fazli, S. (2011). Bristle-sensors – low-cost flexible passive dry EEG electrodes for neurofeedback and BCI applications. *Journal of Neural Engineering*, 8:025008. doi: 10.1088/1741-2560/8/2/025008.
- Guger, C., Daban, S., Sellers, E., Holzner, C., Krausz, G., Carabalona, R., Gramatica, F., and Edlinger, G. (2009). How many people are able to control a P300-based brain-computer interface (BCI)? *Neuroscience Letters*, 462(1):94–98. doi: 10.1016/j.neulet.2009.06.045.
- Guger, C., Krausz, G., Allison, B. Z., and Edlinger, G. (2012). Comparison of dry and gel based electrodes for P300 brain-computer interfaces. *Frontiers in Neuroscience*, 6(60):1–7. doi: 10.3389/fnins.2012.00060.

Bibliography

- Hallez, H., Vanrumste, B., Grech, R., Muscat, J., De Clercq, W., Vergult, A., D'Asseler, Y., Camilleri, K. P., Fabri, S. G., Van Huffel, S., and Lemahieu, I. (2007). Review on solving the forward problem in EEG source analysis. *Journal of NeuroEngineering and Rehabilitation*, 4(1):1–29. doi: 10.1186/1743-0003-4-46.
- Hämäläinen, M., Hari, R., Ilmoniemi, R. J., Knuutila, J., and Lounasmaa, O. V. (1993). Magnetoencephalography – theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics*, 65(2):413. doi: 10.1103/RevModPhys.65.413.
- Harrivel, A., Weissman, D., Noll, D., and Peltier, S. (2013). Monitoring attentional state with fNIRS. *Frontiers in Human Neuroscience*, 7:861. doi: 10.3389/fnhum.2013.00861.
- Hartmann, E., Lachenmayr, B., and Brettel, H. (1979). The peripheral critical flicker frequency. *Vision Research*, 19(9):1019–1023. doi: 10.1016/0042-6989(79)90227-X.
- Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J.-D., Blankertz, B., and Bießmann, F. (2014). On the interpretation of weight vectors of linear models in multivariate neuroimaging. *NeuroImage*, 87:96–110. doi: 10.1016/j.neuroimage.2013.10.067.
- Haufe, S., Treder, M. S., Gugler, M. F., Sagebaum, M., Curio, G., and Blankertz, B. (2011). EEG potentials predict upcoming emergency brakings during simulated driving. *Journal of Neural Engineering*, 8(5):056001. doi: 10.1088/1741-2560/8/5/056001.
- Haynes, J.-D. and Rees, G. (2006). Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience*, 7(7):523–534. doi: 10.1038/nrn1931.
- Herrmann, C. S. (2001). Human EEG responses to 1–100 Hz flicker: Resonance phenomena in visual cortex and their potential correlation to cognitive phenomena. *Experimental Brain Research*, 137(3-4):346–353. doi: 10.1007/s002210100682.
- Hettinger, L. J., Branco, P., Encarnacao, L. M., and Bonato, P. (2003). Neuroadaptive technologies: Applying neuroergonomics to the design of advanced interfaces. *Theoretical Issues in Ergonomics Science*, 4(1-2):220–237. doi: 10.1080/1463922021000020918.
- Hochberg, L. R., Serruya, M. D., Friehs, G. M., Mukand, J. A., Saleh, M., Caplan, A. H., Branner, A., Chen, D., Penn, R. D., and Donoghue, J. P. (2006). Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature*, 442(7099):164–171. doi: 10.1038/nature04970.
- Höhne, J., Holz, E., Staiger-Sälzer, P., Müller, K.-R., Kübler, A., and Tangermann, M. (2014). Motor imagery for severely motor-impaired patients: Evidence for brain-computer interfacing as superior control solution. *PLoS ONE*, 9(8):1–11. doi: 10.1371/journal.pone.0104854.
- Höhne, J. and Tangermann, M. (2014). Towards user-friendly spelling with an auditory brain-computer interface: The CharStreamer paradigm. *PLoS ONE*, 9(6):1–12. doi: 10.1371/journal.pone.0098322.

- Hubel, D. H. and Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology*, 160(1):106–154. doi: 10.1113/j-physiol.1962.sp006837.
- Huynh-Thu, Q., Le Callet, P., and Barkowsky, M. (2010). Video quality assessment: From 2D to 3D - Challenges and future trends. pages 4025–4028. IEEE. doi: 10.1109/ICIP.2010.5650571.
- Hwang, H.-J., Kim, S., Choi, S., and Im, C.-H. (2013). EEG-based brain-computer interfaces: A thorough literature survey. *International Journal of Human-Computer Interaction*, 29(12):814–826. doi: 10.1080/10447318.2013.780869.
- Hwang, H.-J., Lim, J.-H., Jung, Y.-J., C, H., Lee, S. W., and Im, C.-H. (2012). Development of an SSVEP-based BCI spelling system adopting a QWERTY-style LED keyboard. *Journal of Neuroscience Methods*, 208(1):59 – 65. doi: 10.1016/j.jneumeth.2012.04.011.
- Hyvärinen, A., Karhunen, J., and Oja, E. (2001). *Independent Component Analysis*, volume 46. John Wiley & Sons, New York, USA.
- Hyvärinen, A. and Oja, E. (2000). Independent component analysis: Algorithms and applications. *Neural Networks*, 13(4–5):411 – 430. doi: 10.1016/S0893-6080(00)00026-5.
- Jangraw, D. C., Wang, J., Lance, B. J., Chang, S.-F., and Sajda, P. (2014). Neurally and ocularly informed graph-based models for searching 3D environments. *Journal of Neural Engineering*, 11(4):046003. doi: 10.1088/1741-2560/11/4/046003.
- Jayaram, V., Alamgir, M., Altun, Y., Schölkopf, B., and Grosse-Wentrup, M. (2016). Transfer learning in brain-computer interfaces. *IEEE Computational Intelligence Magazine*, 11(1):20–31. doi: 10.1109/MCI.2015.2501545.
- Jeon, Y.-W. and Polich, J. (2001). P3a from a passive visual stimulus task. *Clinical Neurophysiology*, 112(12):2202–2208. doi: 10.1016/S1388-2457(01)00663-0.
- Jung, Y. J., Kim, D., Sohn, H., Lee, S.-i., Park, H. W., and Ro, Y. M. (2013). Subjective and objective measurements of visual fatigue induced by excessive disparities in stereoscopic images. In Woods, A. J., Holliman, N. S., and Favalora, G. E., editors, *Proc. SPIE 8648, Stereoscopic Displays and Applications XXIV*, page 86480M, Burlingame, CA, USA. doi: 10.1117/12.2005818.
- Kamienkowski, J. E., Ison, M. J., Quiroga, R. Q., and Sigman, M. (2012). Fixation-related potentials in visual search: A combined EEG and eye tracking study. *Journal of Vision*, 12(7):4. <http://jov.arvojournals.org/article.aspx?articleid=2297282>.
- Kandel, E. R., Schwartz, J. H., and Jessell, T. M., editors (2000). *Principles of Neural Science*. McGraw-Hill, Health Professions Division, New York, USA, 4th edition.

Bibliography

- Kanoh, S., Miyamoto, K., and Yoshinobu, T. (2008). A brain-computer interface (BCI) system based on auditory stream segregation. In *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 642–645. IEEE. doi: 10.1109/IEMBS.2008.4649234.
- Kaufmann, T., Schulz, S. M., Grünzinger, C., and Kübler, A. (2011). Flashing characters with famous faces improves ERP-based brain-computer interface performance. *Journal of Neural Engineering*, 8(5):056016. doi: 10.1088/1741-2560/8/5/056016.
- Kaunitz, L. N., Kamienkowski, J. E., Varatharajah, A., Sigman, M., Quiroga, R. Q., and Ison, M. J. (2014). Looking for a face in the crowd: Fixation-related potentials in an eye-movement visual search task. *NeuroImage*, 89:297–305. doi: 10.1016/j.neuroimage.2013.12.006.
- Kauppi, J.-P., Kandemir, M., Saarinen, V.-M., Hirvenkari, L., Parkkonen, L., Klami, A., Hari, R., and Kaski, S. (2015). Towards brain-activity-controlled information retrieval: Decoding image relevance from MEG signals. *NeuroImage*, 112:288–298. doi: 10.1016/j.neuroimage.2014.12.079.
- Kim, D., Jung, Y. J., Han, Y., Choi, J., Kim, E., Jeong, B. S., Ro, Y. M., and Park, H. W. (2014). fMRI analysis of excessive binocular disparity on the human brain. *International Journal of Imaging Systems and Technology*, 24(1):94–102. doi: 10.1002/ima.22083.
- Kim, D., Jung, Y. J., Kim, E., Ro, Y. M., and Park, H. W. (2011). Human brain response to visual fatigue caused by stereoscopic depth perception. In *17th International Conference on Digital Signal Processing (DSP)*, pages 1–5. doi: 10.1109/ICDSP2011.6004997.
- Kim, Y.-J. and Lee, E. C. (2011). EEG based comparative measurement of visual fatigue caused by 2D and 3D displays. In Stephanidis, C., editor, *HCI International 2011 – Posters' Extended Abstracts*, number 174 in Communications in Computer and Information Science, pages 289–292. Springer, Berlin, Heidelberg, Germany. doi: 10.1007/978-3-642-22095-1_59.
- Kindermans, P.-J., Tangermann, M., Müller, K.-R., and Schrauwen, B. (2014). Integrating dynamic stopping, transfer learning and language models in an adaptive zero-training ERP speller. *Journal of Neural Engineering*, 11(3):035005. doi: 10.1088/1741-2560/11/3/035005.
- Kliegl, R., Dambacher, M., Dimigen, O., Jacobs, A. M., and Sommer, W. (2012). Eye movements and brain electric potentials during reading. *Psychological Research*, 76(2):145–158. doi: 10.1007/s00426-011-0376-x.
- Knöpfel, T. (2012). Genetically encoded optical indicators for the analysis of neuronal circuits. *Nature Reviews Neuroscience*, 13(10):687–700. doi: 10.1038/nrn3293.
- Koch, C. (2004). *Biophysics of computation: Information processing in single neurons*. Oxford University Press, Oxford, UK; New York, USA.
- Koch, C. and Segev, I. (1998). *Methods in neuronal modeling: From ions to networks*. MIT Press, Cambridge, MA, USA.

- Kohlmorgen, J., Dornhege, G., Braun, M., Blankertz, B., Müller, K.-R., Curio, G., Hagemann, K., Bruns, A., Schrauf, M., Kincses, W., et al. (2007). Improving human performance in a real operating environment through real-time mental workload detection. In Dornhege, G. R., del R. Millán, J., Hinterberger, T., McFarland, D., and Müller, K.-R., editors, *Toward Brain-Computer Interfacing*. MIT Press, Cambridge, MA, USA.
- Koles, Z. J., Lazar, M. S., and Zhou, S. Z. (1990). Spatial patterns underlying population differences in the background EEG. *Brain Topography*, 2(4):275–284. doi: 10.1007/BF01129656.
- Kollmorgen, S., Nortmann, N., Schröder, S., and König, P. (2010). Influence of low-level stimulus features, task dependent factors, and spatial biases on overt visual attention. *PLoS Computational Biology*, 6(5):e1000791. doi: 10.1371/journal.pcbi.1000791.
- Kooi, F. L. and Toet, A. (2004). Visual comfort of binocular and 3D displays. *Displays*, 25(2–3):99–108. doi: 10.1016/j.displa.2004.07.004.
- Kornrumpf, B., Niefind, F., Sommer, W., and Dimigen, O. (2016). Neural correlates of word recognition: A systematic comparison of natural reading and rapid serial visual presentation. *Journal of Cognitive Neuroscience*, 28(9):1374–1391. doi: 10.1162/jocn_a_00977.
- Koyamada, S., Shikauchi, Y., Nakae, K., Koyama, M., and Ishii, S. (2015). Deep learning of fMRI big data: A novel approach to subject-transfer decoding. *arXiv preprint arXiv:1502.00093*. <https://arxiv.org/abs/1502.00093>.
- Kübler, A., Nijboer, F., Mellinger, J., Vaughan, T. M., Pawelzik, H., Schalk, G., McFarland, D. J., Birbaumer, N., and Wolpaw, J. R. (2005). Patients with ALS can use sensorimotor rhythms to operate a brain-computer interface. *Neurology*, 64(10):1775–1777. doi: 10.1212/01.WNL.0000158616.43002.6D.
- Lambooi, M., Ijsselsteijn, W., Bouwhuis, D. G., and Heynderickx, I. (2011a). Evaluation of stereoscopic images: Beyond 2D quality. *IEEE Transactions on Broadcasting*, 57(2):432–444. doi: 10.1109/TBC.2011.2134590.
- Lambooi, M., Ijsselsteijn, W. A., and Heynderickx, I. (2011b). Visual discomfort of 3D TV: Assessment methods and modeling. *Displays*, 32(4):209–218. doi: 10.1016/j.displa.2011.05.012.
- Lambooi, M. T. M., Ijsselsteijn, W. A., and Heynderickx, I. (2007). Visual discomfort in stereoscopic displays: A review. In Woods, A. J., Dodgson, N. A., Merritt, J. O., Bolas, M. T., and McDowall, I. E., editors, *Proc. SPIE 6490, Stereoscopic Displays and Virtual Reality Systems XIV*, page 64900I, San Jose, CA, USA. doi: 10.1117/12.705527.
- Lebedev, M. A. and Nicolelis, M. A. L. (2006). Brain-machine interfaces: Past, present and future. *Trends in Neurosciences*, 29(9):536–546. doi: 10.1016/j.tins.2006.07.004.
- Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2):365–411. doi: 10.1016/S0047-259X(03)00096-4.

Bibliography

- Lemm, S., Blankertz, B., Dickhaus, T., and Müller, K.-R. (2011). Introduction to machine learning for brain imaging. *NeuroImage*, 56(2):387–399. doi: 10.1016/j.neuroimage.2010.11.004.
- Lev, J. et al. (1949). The point biserial coefficient of correlation. *The Annals of Mathematical Statistics*, 20(1):125–126. doi: 10.1214/aoms/1177730103.
- Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, 49(2B):467–477. doi: 10.1121/1.1912375.
- Li, H.-C. O., Seo, J., Kham, K., and Lee, S. (2008). Measurement of 3D visual fatigue using event-related potential (ERP): 3D oddball paradigm. In *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video, 2008*, pages 213–216. doi: 10.1109/3DTV.2008.4547846.
- Lindemann, L., Wenger, S., and Magnor, M. (2011). Evaluation of video artifact perception using event-related potentials. In *Proceedings of the ACM SIGGRAPH Symposium on Applied Perception in Graphics and Visualization*, pages 53–58. Association for Computing Machinery. doi: 10.1145/2077451.2077461.
- Liu, Y., Zhou, Z., and Hu, D. (2011). Gaze independent brain-computer speller with covert visual search tasks. *Clinical Neurophysiology*, 122(6):1127–1136. doi: 10.1016/j.clinph.2010.10.049.
- Logothetis, N. K. (2008). What we can do and what we cannot do with fMRI. *Nature*, 453(7197):869–878. doi: 10.1038/nature06976.
- Looney, D., Kidmose, P., and Mandic, D. P. (2014). Ear-EEG: User-centered and wearable BCI. In *Brain-Computer Interface Research*, pages 41–50. Springer, Berlin, Heidelberg, Germany. doi: 10.1007/978-3-642-54707-2_5.
- Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., and Arnaldi, B. (2007). A review of classification algorithms for EEG-based brain-computer interfaces. *Journal of Neural Engineering*, 4(2):R1. doi: 10.1088/1741-2560/4/2/R01.
- Lu, S., Guan, C., and Zhang, H. (2009). Unsupervised brain computer interface based on inter-subject information and online adaptation. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 17(2):135–145. doi: 10.1109/TNSRE.2009.2015197.
- Lyskov, E., Ponomarev, V., Sandström, M., Hansson Mild, K., and Medvedev, S. (1998). Steady-state visual evoked potentials to computer monitor flicker. *International Journal of Psychophysiology*, 28(3):285–290. doi: 10.1016/S0167-8760(97)00074-3.
- Mak, J. N. and Wolpaw, J. R. (2009). Clinical applications of brain-computer interfaces: current state and future prospects. *IEEE Reviews in Biomedical Engineering*, 2:187. doi: 10.1109/RBME.2009.2035356.

- Makeig, S., Bell, A. J., Jung, T.-P., Sejnowski, T. J., et al. (1996a). Independent component analysis of electroencephalographic data. *Advances in Neural Information Processing Systems*, pages 145–151. <https://papers.nips.cc/paper/1091-independent-component-analysis-of-electroencephalographic-data.pdf>.
- Makeig, S., Jung, T.-P., and Sejnowski, T. J. (1996b). Using feedforward neural networks to monitor alertness from changes in EEG correlation and coherence. *Advances in Neural Information Processing Systems*, pages 931–937.
- Malmivuo, J. A. and Suihko, V. E. (2004). Effect of skull resistivity on the spatial resolutions of EEG and MEG. *IEEE Transactions on Biomedical Engineering*, 51(7):1276–1280. doi: 10.1109/TBME.2004.827255.
- Manyakov, N. V., Chumerin, N., Combaz, A., and Van Hulle, M. M. (2011). Comparison of classification methods for P300 brain-computer interface on disabled subjects. *Computational Intelligence and Neuroscience*, 2011:2. doi: 10.1155/2011/519868.
- Marathe, A. R., Ries, A. J., and McDowell, K. (2013). A novel method for single-trial classification in the face of temporal variability. In Schmorow, D. D. and Fidopiastis, C. M., editors, *Foundations of Augmented Cognition*, number 8027 in Lecture Notes in Computer Science, pages 345–352. Springer, Berlin, Heidelberg, Germany.
- Marathe, A. R., Ries, A. J., and McDowell, K. (2014). Sliding HDCA: Single-trial EEG classification to overcome and quantify temporal variability. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 22(2):201–211. doi: 10.1109/TNSRE.2014.2304884.
- Marron, J. S., Ramsay, J. O., Sangalli, L. M., Srivastava, A., et al. (2015). Functional data analysis of amplitude and phase variation. *Statistical Science*, 30(4):468–484. doi: 10.1214/15-STS524.
- McKinney, W. (2010). Data structures for statistical computing in Python. In van der Walt, S. and Millman, J., editors, *Proceedings of the 9th Python in Science Conference*, pages 51–56. <http://204.236.236.243/proceedings/scipy2010/pdfs/mckinney.pdf>.
- McKinney, W. (2011). Pandas: A foundational Python library for data analysis and statistics. *Python for High Performance and Scientific Computing*, pages 1–9. http://www.dlr.de/sc/Portaldata/15/Resources/dokumente/pyhpc2011/submissions/pyhpc2011_submission_9.pdf.
- Mehta, K. and Kliewer, J. (2015). Directed information measures for assessing perceived audio quality using EEG. In *2015 49th Asilomar Conference on Signals, Systems and Computers*, pages 123–127. IEEE. doi: 10.1109/ACSSC.2015.7421096.
- Mehta, R. K. and Parasuraman, R. (2013). Neuroergonomics: A review of applications to physical and cognitive work. *Frontiers in Human Neuroscience*, 7:889. doi: 10.3389/fnhum.2013.00889.

Bibliography

- Meng, J., Merino, L., Robbins, K., and Huang, Y. (2012). Classification of EEG recordings without perfectly time-locked events. In *2012 IEEE Statistical Signal Processing Workshop (SSP)*, pages 444–447. doi: 10.1109/SSP.2012.6319727.
- Millán, J. d. R. and Mouriño, J. (2003). Asynchronous BCI and local neural classifiers: An overview of the adaptive brain interface project. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 11(2):159–161. doi: 10.1109/TNSRE.2003.814435.
- Millán, J. d. R., Mouriño, J., Franzé, M., Cincotti, F., Varsta, M., Heikkonen, J., and Babiloni, F. (2002). A local neural classifier for the recognition of EEG patterns associated to mental tasks. *Neural Networks, IEEE Transactions on*, 13(3):678–686. doi: 10.1109/TNN.2002.1000132.
- Millán, J. d. R., Rupp, R., Müller-Putz, G., Murray-Smith, R., Giugliemma, C., Tangermann, M., Vidaurre, C., Cincotti, F., Kübler, A., Leeb, R., Neuper, C., Müller, K.-R., and Mattia, D. (2010). Combining brain–computer interfaces and assistive technologies: State-of-the-art and challenges. *Frontiers in Neuroscience*, 4:161. doi: 10.3389/fnins.2010.00161.
- Miller, K. J., denNijs, M., Shenoy, P., Miller, J. W., Rao, R. P. N., and Ojemann, J. G. (2007). Real-time functional brain mapping using electrocortigraphy. *NeuroImage*, 37(2):504–507. doi: 10.1016/j.neuroimage.2007.05.029.
- Mühl, C., Jeunet, C., and Lotte, F. (2014). EEG-based workload estimation across affective contexts. *Frontiers in Neuroscience*, 8:114. doi: 10.3389/fnins.2014.00114.
- Mullen, T. R., Kothe, C. A. E., Chi, Y. M., Ojeda, A., Kerth, T., Makeig, S., Jung, T.-P., and Cauwenberghs, G. (2015). Real-time neuroimaging and cognitive monitoring using wearable dry EEG. *IEEE Transactions on Biomedical Engineering*, 62(11):2553–2567. doi: 10.1109/TBME.2015.2481482.
- Müller, J. S., Vidaurre, C., Schreuder, M., Meinecke, F., von Büna, P., and Müller, K.-R. (2017). A mathematical model for the two-learners problem. *Journal of Neural Engineering*. doi: 10.1088/1741-2552/aa620b.
- Müller, K.-R., Anderson, C. W., and Birch, G. E. (2003). Linear and nonlinear methods for brain-computer interfaces. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 11(2):165–169. doi: 10.1109/TNSRE.2003.814484.
- Müller, K.-R., Mika, S., Rätsch, G., Tsuda, K., and Schölkopf, B. (2001). An introduction to kernel-based learning algorithms. *IEEE Transactions on Neural Networks*, 12(2):181–201. doi: 10.1109/72.914517.
- Müller, K.-R., Tangermann, M., Dornhege, G., Krauledat, M., Curio, G., and Blankertz, B. (2008). Machine learning for real-time single-trial EEG-analysis: From brain–computer interfacing to mental state monitoring. *Journal of Neuroscience Methods*, 167(1):82–90. doi: 10.1016/j.jneumeth.2007.09.022.

- Mun, S., Park, M.-C., Park, S., and Whang, M. (2012). SSVEP and ERP measurement of cognitive fatigue caused by stereoscopic 3D. *Neuroscience Letters*, 525(2):89–94. doi: 10.1016/j.neulet.2012.07.049.
- Musallam, S., Corneil, B. D., Greger, B., Scherberger, H., and Andersen, R. A. (2004). Cognitive control signals for neural prosthetics. *Science*, 305(5681):258–262. doi: 10.1126/science.1097938.
- Mustafa, M., Lindemann, L., and Magnor, M. (2012). EEG analysis of implicit human visual perception. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 513–516, Austin, TX, USA. Association for Computing Machinery. doi: 10.1145/2207676.2207746.
- Nicolas-Alonso, L. F. and Gomez-Gil, J. (2012). Brain computer interfaces, a review. *Sensors*, 12(2):1211–1279. doi: 10.3390/s120201211.
- Nicolelis, M. A. L. (2001). Actions from thoughts. *Nature*, 409(6818):403–407. doi: 10.1038/35053191.
- Nikulin, V. V., Kegeles, J., and Curio, G. (2010). Miniaturized electroencephalographic scalp electrode for optimal wearing comfort. *Clinical Neurophysiology*, 121(7):1007–1014. doi: 10.1016/j.clinph.2010.02.008.
- Norton, J. J. S., Lee, D. S., Lee, J. W., Lee, W., Kwon, O., Won, P., Jung, S.-Y., Cheng, H., Jeong, J.-W., Akce, A., et al. (2015). Soft, curved electrode systems capable of integration on the auricle as a persistent brain–computer interface. *Proceedings of the National Academy of Sciences*, 112(13):3920–3925. doi: 10.1073/pnas.1424875112.
- Parasuraman, R. (2003). Neuroergonomics: Research and practice. *Theoretical Issues in Ergonomics Science*, 4(1-2):5–20. doi: 10.1080/14639220210199753.
- Parasuraman, R., Christensen, J., and Grafton, S. (2012). Neuroergonomics: The brain in action and at work. *NeuroImage*, 59(1):1–3. doi: 10.1016/j.neuroimage.2011.08.011.
- Parasuraman, R. and Rizzo, M. (2008). *Neuroergonomics: The brain at work*. Oxford University Press, Oxford, UK; New York, USA.
- Parra, L. C., Spence, C. D., Gerson, A. D., and Sajda, P. (2003). Response error correction – a demonstration of improved human-machine performance using real-time EEG monitoring. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 11(2):173–177. doi: 10.1109/TNSRE.2003.814446.
- Parra, L. C., Spence, C. D., Gerson, A. D., and Sajda, P. (2005). Recipes for the linear analysis of EEG. *NeuroImage*, 28(2):326–341. doi: 10.1016/j.neuroimage.2005.05.032.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher,

Bibliography

- M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830. <http://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf>.
- Penrose, R. (1955). A generalized inverse for matrices. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 51, pages 406–413, Cambridge, UK. Cambridge University Press. doi: 10.1017/S0305004100030401.
- Petersen, K., Hansen, L. K., Kolenda, T., Rostrup, E., and Strother, S. (2000). On the independent components of functional neuroimages. In *Third International Conference on Independent Component Analysis and Blind Source Separation*, pages 615–620. <http://research.ics.aalto.fi/events/ica2000/proceedings/0615.pdf>.
- Pfurtscheller, G., Brunner, C., Schlögl, A., and Da Silva, F. H. L. (2006). Mu rhythm (de) synchronization and EEG single-trial classification of different motor imagery tasks. *NeuroImage*, 31(1):153–159. doi: 10.1016/j.neuroimage.2005.12.003.
- Picton, T. W. (1992). The P300 wave of the human event-related potential. *Journal of Clinical Neurophysiology*, 9(4):456–479.
- Piwek, L., Ellis, D. A., Andrews, S., and Joinson, A. (2016). The rise of consumer health wearables: Promises and barriers. *PLoS Medicine*, 13(2):e1001953. doi: 10.1371/journal.pmed.1001953.
- Plöchl, M., Ossandón, J. P., and König, P. (2012). Combining EEG and eye tracking: Identification, characterization, and correction of eye movement artifacts in electroencephalographic data. *Frontiers in Human Neuroscience*, 6:278. doi: 10.3389/fnhum.2012.00278.
- Pohlmeyer, E. A., Wang, J., Jangraw, D. C., Lou, B., Chang, S.-F., and Sajda, P. (2011). Closing the loop in cortically-coupled computer vision: A brain-computer interface for searching image databases. *Journal of Neural Engineering*, 8(3):036025. doi: 10.1088/1741-2560/8/3/036025.
- Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, 118(10):2128–2148. doi: 10.1016/j.clinph.2007.04.019.
- Pope, A. T., Bogart, E. H., and Bartolome, D. S. (1995). Biocybernetic system evaluates indices of operator engagement in automated task. *Biological Psychology*, 40(1-2):187 – 195. doi: 10.1016/0301-0511(95)05116-3.
- Popescu, F., Fazli, S., Badower, Y., Blankertz, B., and Müller, K.-R. (2007). Single trial classification of motor imagination using 6 dry EEG electrodes. *PLoS ONE*, 2(7):1–5. doi: 10.1371/journal.pone.0000637.
- Porbadnigk, A. K., Antons, J., Blankertz, B., Treder, M. S., Schleicher, R., Möller, S., and Curio, G. (2010). Using ERPs for assessing the (sub) conscious perception of noise. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, pages 2690–2693. IEEE. doi: 10.1109/IEMBS.2010.5626549.

- Porbadnigk, A. K., Görnitz, N., Sannelli, C., Binder, A., Braun, M., Kloft, M., and Müller, K.-R. (2015). Extracting latent brain states – Towards true labels in cognitive neuroscience experiments. *NeuroImage*, 120:225–253. doi: 10.1016/j.neuroimage.2015.05.078.
- Porbadnigk, A. K., Scholler, S., Blankertz, B., Ritz, A., Born, M., Scholl, R., Müller, K.-R., Curio, G., and Treder, M. S. (2011). Revealing the neural response to imperceptible peripheral flicker with machine learning. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC*, pages 3692–3695. doi: 10.1109/IEMBS.2011.6090625.
- Porbadnigk, A. K., Treder, M. S., Blankertz, B., Antons, J.-N., Schleicher, R., Möller, S., Curio, G., and Müller, K.-R. (2013). Single-trial analysis of the neural correlates of speech quality perception. *Journal of Neural Engineering*, 10(5):056003. doi: 10.1088/1741-2560/10/5/056003.
- Ramoser, H., Müller-Gerking, J., and Pfurtscheller, G. (2000). Optimal spatial filtering of single trial EEG during imagined hand movement. *IEEE Transactions on Rehabilitation Engineering*, 8(4):441–446. doi: 10.1109/86.895946.
- Ramsay, J. and Silverman, B. W. (2005). *Functional Data Analysis*. Springer, New York, USA. doi: 10.1007/b98888.
- Samek, W., Kawanabe, M., and Müller, K.-R. (2014). Divergence-based framework for common spatial patterns algorithms. *IEEE Reviews in Biomedical Engineering*, 7:50–72. doi: 10.1109/RBME.2013.2290621.
- Schäfer, J. and Strimmer, K. (2005). A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical Applications in Genetics and Molecular Biology*, 4(1). doi: 10.2202/1544-6115.1175.
- Schalk, G., Wolpaw, J. R., McFarland, D. J., and Pfurtscheller, G. (2000). EEG-based communication: Presence of an error potential. *Clinical Neurophysiology*, 111(12):2138–2144. doi: 10.1016/S1388-2457(00)00457-0.
- Schmidt, E. M. (1980). Single neuron recording from motor cortex as a possible source of signals for control of external devices. *Annals of Biomedical Engineering*, 8(4):339–349. doi: 10.1007/BF02363437.
- Schmidt, N. M., Blankertz, B., and Treder, M. S. (2012). Online detection of error-related potentials boosts the performance of mental typewriters. *BMC Neuroscience*, 13(1):1. doi: 10.1186/1471-2202-13-19.
- Scholler, S., Bosse, S., Treder, M. S., Blankertz, B., Curio, G., Müller, K.-R., and Wiegand, T. (2012). Toward a direct measure of video quality perception using EEG. *IEEE Transactions on Image Processing*, 21(5):2619–2629. doi: 10.1109/TIP.2012.2187672.
- Schreuder, M., Rost, T., and Tangermann, M. (2011). Listen, you are writing! Speeding up online spelling with a dynamic auditory BCI. *Frontiers in Neuroscience*, 5:112. doi: 10.3389/fnins.2011.00112.

Bibliography

- Schultze-Kraft, M., Birman, D., Rusconi, M., Allefeld, C., Gorgen, K., Dahne, S., Blankertz, B., and Haynes, J.-D. (2016a). The point of no return in vetoing self-initiated movements. *Proceedings of the National Academy of Sciences*, 113(4):1080–1085. doi: 10.1073/pnas.1513569112.
- Schultze-Kraft, M., Dahne, S., Gugler, M., Curio, G., and Blankertz, B. (2016b). Unsupervised classification of operator workload from brain signals. *Journal of Neural Engineering*, 13(3):036008. doi: 10.1088/1741-2560/13/3/036008.
- Schultze-Kraft, R., Gorgen, K., Wenzel, M. A., Haynes, J. D., and Blankertz, B. (2013). Cooperating brains: Joint control of a dual-BCI. In *Proceedings of the 5th International Brain-Computer Interface Meeting*, Asilomar, USA.
- Schwartz, A. B., Cui, X. T., Weber, D., and Moran, D. W. (2006). Brain-controlled interfaces: Movement restoration with neural prosthetics. *Neuron*, 52(1):205 – 220. doi: 10.1016/j.neuron.2006.09.019.
- Sellers, E. W. and Donchin, E. (2006). A P300-based brain–computer interface: Initial tests by ALS patients. *Clinical Neurophysiology*, 117(3):538–548. doi: 10.1016/j.clinph.2005.06.027.
- Seoane, L. F., Gabler, S., and Blankertz, B. (2015). Images from the mind: BCI image evolution based on rapid serial visual presentation of polygon primitives. *Brain-Computer Interfaces*, 2(1):40–56. doi: 10.1080/2326263X.2015.1060819.
- Shady, S., MacLeod, D. I. A., and Fisher, H. S. (2004). Adaptation from invisible flicker. *Proceedings of the National Academy of Sciences of the United States of America*, 101(14):5170–5173. doi: 10.1073/pnas.0303452101.
- Shen, K.-Q., Li, X.-P., Ong, C.-J., Shao, S.-Y., and Wilder-Smith, E. P. V. (2008). EEG-based mental fatigue measurement using multi-class support vector machines with confidence estimate. *Clinical Neurophysiology*, 119(7):1524–1533. doi: 10.1016/j.clinph.2008.03.012.
- Sherrington, C. S. (1904). On binocular flicker and the correlation of activity of ‘corresponding’ retinal points. *British Journal of Psychology, 1904-1920*, 1(1):26–60. doi: 10.1111/j.2044-8295.1904.tb00148.x.
- Silver, N. C. and Dunlap, W. P. (1987). Averaging correlation coefficients: Should Fisher’s z transformation be used? *Journal of Applied Psychology*, 72(1):146. doi: 10.1037/0021-9010.72.1.146.
- Sitaram, R., Ros, T., Stoeckel, L., Haller, S., Scharnowski, F., Lewis-Peacock, J., Weiskopf, N., Blefari, M. L., Rana, M., Oblak, E., Birbaumer, N., and Sulzer, J. (2017). Closed-loop brain training: The science of neurofeedback. *Nature Reviews Neuroscience*, 18(2):86–100. doi: 10.1038/nrn.2016.164.

- Sitaram, R., Zhang, H., Guan, C., Thulasidas, M., Hoshi, Y., Ishikawa, A., Shimizu, K., and Birbaumer, N. (2007). Temporal classification of multichannel near-infrared spectroscopy signals of motor imagery for developing a brain-computer interface. *NeuroImage*, 34(4):1416–1427. doi: 10.1016/j.neuroimage.2006.11.005.
- Squires, N. K., Squires, K. C., and Hillyard, S. A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalography and Clinical Neurophysiology*, 38(4):387–401. doi: 10.1016/0013-4694(75)90263-1.
- Stikic, M., Johnson, R. R., Levendowski, D. J., Popovic, D. P., Olmstead, R. E., and Berka, C. (2011). EEG-derived estimators of present and future cognitive performance. *Frontiers in Human Neuroscience*, 5:70. doi: 10.3389/fnhum.2011.00070.
- Stopczynski, A., Stahlhut, C., Larsen, J. E., Petersen, M. K., and Hansen, L. K. (2014). The smartphone brain scanner: A portable real-time neuroimaging system. *PLoS ONE*, 9(2):e86733. doi: 10.1371/journal.pone.0086733.
- Stosiek, C., Garaschuk, O., Holthoff, K., and Konnerth, A. (2003). In vivo two-photon calcium imaging of neuronal networks. *Proceedings of the National Academy of Sciences*, 100(12):7319–7324. doi: 10.1073/pnas.1232232100.
- Sutter, E. E. (1984). The visual evoked response as a communication channel. In *Proceedings of the IEEE Symposium on Biosensors*, pages 95–100.
- Sutton, S., Braren, M., Zubin, J., and John, E. R. (1965). Evoked-potential correlates of stimulus uncertainty. *Science*, 150(3700):1187–1188. doi: 10.1126/science.150.3700.1187.
- Tam, W. J., Speranza, F., Yano, S., Shimono, K., and Ono, H. (2011). Stereoscopic 3D-TV: Visual comfort. *IEEE Transactions on Broadcasting*, 57(2):335–346. doi: 10.1109/TBC.2011.2125070.
- Tatler, B. W., Hayhoe, M. M., Land, M. F., and Ballard, D. H. (2011). Eye guidance in natural vision: Reinterpreting salience. *Journal of Vision*, 11(5):5. doi: 10.1167/11.5.5.
- Telenczuk, B., Baker, S. N., Herz, A. V. M., and Curio, G. (2011). High-frequency EEG co-varies with spike burst patterns detected in cortical neurons. *Journal of Neurophysiology*, 105(6):2951–2959. doi: 10.1152/jn.00327.2010.
- Thickbroom, G., Knezevič, W., Carroll, W., and Mastaglia, F. (1991). Saccade onset and offset lambda waves: Relation to pattern movement visually evoked potentials. *Brain Research*, 551(1–2):150–156. doi: 10.1016/0006-8993(91)90927-N.
- Thickbroom, G. and Mastaglia, F. (1985). Presaccadic ‘spike’ potential: Investigation of topography and source. *Brain Research*, 339(2):271 – 280. doi: 10.1016/0006-8993(85)90092-7.
- Thielen, J., van den Broek, P., Farquhar, J., and Desain, P. (2015). Broad-band visually evoked potentials: Re(con)volution in brain-computer interfacing. *PLoS ONE*, 10(7):1–22. doi: 10.1371/journal.pone.0133797.

Bibliography

- Ting, S., Tan, T., West, G., Squelch, A., and Foster, J. (2011). Quantitative assessment of 2D versus 3D visualisation modalities. In *2011 IEEE Visual Communications and Image Processing (VCIP)*, pages 1–4. doi: 10.1109/VCIP.2011.6115908.
- Toman, J. (1941). Flicker potentials and the alpha rhythm in man. *Journal of Neurophysiology*, 4(1):51.
- Tomioka, R. and Müller, K.-R. (2010). A regularized discriminative framework for EEG analysis with application to brain–computer interface. *NeuroImage*, 49(1):415–432. doi: 10.1016/j.neuroimage.2009.07.045.
- Traditional (1842). Die Gedanken sind frei. In Fallersleben, H. v. and Richter, E., editors, *Schlesische Volkslieder mit Melodien*, page 307. Breitkopf und Härtel, Leipzig.
- Trappenberg, T. (2009). *Fundamentals of Computational Neuroscience*. Oxford University Press, Oxford, UK; New York, USA.
- Treder, M. S. and Blankertz, B. (2010). (C)overt attention and visual speller design in an ERP-based brain-computer interface. *Behavioral and Brain Functions*, 6(1):28. doi: 10.1186/1744-9081-6-28.
- Treder, M. S., Purwins, H., Miklody, D., Sturm, I., and Blankertz, B. (2014). Decoding auditory attention to instruments in polyphonic music using single-trial EEG classification. *Journal of Neural Engineering*, 11(2):026009. doi: 10.1088/1741-2560/11/2/026009.
- Treder, M. S., Schmidt, N. M., and Blankertz, B. (2011). Gaze-independent brain–computer interfaces based on covert attention and feature attention. *Journal of Neural Engineering*, 8(6):066003. doi: 10.1088/1741-2560/8/6/066003.
- Tsoneva, T., Garcia-Molina, G., and Desain, P. (2015). Neural dynamics during repetitive visual stimulation. *Journal of Neural Engineering*, 12(6):066017. doi: 10.1088/1741-2560/12/6/066017.
- Tulyakov, S., Jaeger, S., Govindaraju, V., and Doermann, D. (2008). Review of classifier combination methods. In Marinai, S. and Fujisawa, H., editors, *Machine Learning in Document Analysis and Recognition*, number 90 in Studies in Computational Intelligence, pages 361–386. Springer, Berlin, Heidelberg, Germany. doi: 10.1007/978-3-540-76280-5_14.
- Ukai, K. and Howarth, P. A. (2008). Visual fatigue caused by viewing stereoscopic motion images: Background, theories, and observations. *Displays*, 29(2):106–116. doi: 10.1016/j.displa.2007.09.004.
- Ulbert, I., Halgren, E., Heit, G., and Karmos, G. (2001). Multiple microelectrode-recording system for human intracortical applications. *Journal of Neuroscience Methods*, 106(1):69–79. doi: 10.1016/S0165-0270(01)00330-2.
- Urigüen, J. A. and Garcia-Zapirain, B. (2015). EEG artifact removal – state-of-the-art and guidelines. *Journal of Neural Engineering*, 12(3):031001. doi: 10.1088/1741-2560/12/3/031001.

- Ušćumlić, M. and Blankertz, B. (2016). Active visual search in non-stationary scenes: Coping with temporal variability and uncertainty. *Journal of Neural Engineering*, 13(1):016015. doi: 10.1088/1741-2560/13/1/016015.
- Ušćumlić, M., Chavarriaga, R., and Millán, J. d. R. (2013). An iterative framework for EEG-based image search: Robust retrieval with weak classifiers. *PLoS ONE*, 8(8):e72018. doi: 10.1371/journal.pone.0072018.
- van der Waal, M., Severens, M., Geuze, J., and Desain, P. (2012). Introducing the tactile speller: An ERP-based brain-computer interface for communication. *Journal of Neural Engineering*, 9(4):045002. doi: 10.1088/1741-2560/9/4/045002.
- van Gerven, M., Farquhar, J., Schaefer, R., Vlek, R., Geuze, J., Nijholt, A., Ramsey, N., Haselager, P., Vuurpijl, L., Gielen, S., and Desain, P. (2009). The brain-computer interface cycle. *Journal of Neural Engineering*, 6(4):041001. doi: 10.1088/1741-2560/6/4/041001.
- Vansteensel, M. J., Pels, E. G. M., Bleichner, M. G., Branco, M. P., Denison, T., Freudenburg, Z. V., Gosselaar, P., Leinders, S., Ottens, T. H., Van Den Boom, M. A., Van Rijen, P. C., Aarnoutse, E. J., and Ramsey, N. F. (2016). Fully implanted brain-computer interface in a locked-in patient with ALS. *New England Journal of Medicine*, 375(21):2060–2066. doi: 10.1056/NEJMoa1608085.
- Velliste, M., Perel, S., Spalding, M. C., Whitford, A. S., and Schwartz, A. B. (2008). Cortical control of a prosthetic arm for self-feeding. *Nature*, 453(7198):1098–1101. doi: 10.1038/nature06996.
- Venthur, B., Scholler, S., Williamson, J., Dähne, S., Treder, M. S., Kramarek, M. T., Müller, K.-R., and Blankertz, B. (2010). Pyff – a Pythonic framework for feedback applications and stimulus presentation in neuroscience. *Frontiers in Neuroinformatics*, 4:100. doi: 10.3389/fninf.2010.00100.
- Vidal, J. J. (1973). Toward direct brain-computer communication. *Annual Review of Biophysics and Bioengineering*, 2(1):157–180. doi: 10.1146/annurev.bb.02.060173.001105.
- Vidal, J. J. (1977). Real-time detection of brain events in EEG. *Proceedings of the IEEE*, 65(5):633–641. doi: 10.1109/PROC.1977.10542.
- Vidaurre, C., Sannelli, C., Müller, K.-R., and Blankertz, B. (2011a). Co-adaptive calibration to improve BCI efficiency. *Journal of Neural Engineering*, 8(2):025009 (8pp). doi: 10.1088/1741-2560/8/2/025009.
- Vidaurre, C., Sannelli, C., Müller, K.-R., and Blankertz, B. (2011b). Machine-learning based co-adaptive calibration. *Neural Computation*, 23(3):791–816. doi: 10.1162/NECO_a_00089.
- von Büna, P., Meinecke, F. C., Király, F. C., and Müller, K.-R. (2009). Finding stationary subspaces in multivariate time series. *Physical Review Letters*, 103(21):214101. doi: 10.1103/PhysRevLett.103.214101.

Bibliography

- von Lüthmann, A., Herff, C., Heger, D., and Schultz, T. (2015). Toward a wireless open source instrument: Functional near-infrared spectroscopy in mobile neuroergonomics and BCI applications. *Frontiers in Human Neuroscience*, 9. doi: 10.3389/fnhum.2015.00617.
- von Lüthmann, A., Wabnitz, H., Sander, T., and Müller, K.-R. (2016). M3BA: A mobile, modular, multimodal biosignal acquisition architecture for miniaturized EEG-NIRS based hybrid BCI and monitoring. *IEEE Transactions on Biomedical Engineering*. doi: 10.1109/TBME.2016.2594127.
- Wandell, B. A. (1995). *Foundations of Vision*. Sinauer Associates Inc, Sunderland, MA, USA, 1st edition.
- Waskom, M., Botvinnik, O., Hobson, P., Warmenhoven, J., Cole, J. B., Halchenko, Y., Vanderplas, J., Hoyer, S., Villalba, S., Quintero, E., Miles, A., Augspurger, T., Yarkoni, T., Evans, C., Wehner, D., Rocher, L., Megies, T., Coelho, L. P., Ziegler, E., Hoppe, T., Seabold, S., Pascual, S., Cloud, P., Koskinen, M., Hausler, C., Kjemmet, Milajevs, D., Qalieh, A., Allan, D., and Meyer, K. (2015). Seaborn: v0.6.0 (June 2015). doi: 10.5281/zenodo.19108.
- Waytowich, N. R. and Krusienski, D. J. (2015). Spatial decoupling of targets and flashing stimuli for visual brain-computer interfaces. *Journal of Neural Engineering*, 12(3):036006. doi: 10.1088/1741-2560/12/3/036006.
- Wenzel, M. A., Almeida, I., and Blankertz, B. (2016a). The contribution of counting to neural activity evoked by the oddball paradigm. In *Proceedings of the 6th International Brain-Computer Interface Meeting*, Asilomar, USA.
- Wenzel, M. A., Almeida, I., and Blankertz, B. (2016b). Is neural activity detected by ERP-based brain-computer interfaces task specific? *PLoS ONE*, 11(10):1–16. doi: 10.1371/journal.pone.0165556.
- Wenzel, M. A., Bogojeski, M., and Blankertz, B. (2017). Real-time inference of word relevance from electroencephalogram and eye gaze. *Journal of Neural Engineering*. doi: 10.1088/1741-2552/aa7590.
- Wenzel, M. A., Golenia, J.-E., and Blankertz, B. (2016c). Classification of eye fixation related potentials for variable stimulus saliency. *Frontiers in Neuroprosthetics*, 10(23). doi: 10.3389/fnins.2016.00023.
- Wenzel, M. A., Moreira, C., Lungu, I.-A., Bogojeski, M., and Blankertz, B. (2015). Neural responses to abstract and linguistic stimuli with variable recognition latency. In Blankertz, B., Jacucci, G., Gamberini, L., Spagnoli, A., and Freeman, J., editors, *Symbiotic Interaction*, volume 9359 of *Lecture Notes in Computer Science*, pages 172–178. Springer International Publishing. doi: 10.1007/978-3-319-24917-9_19.
- Wenzel, M. A., Schultze-Kraft, R., Meinecke, F. C., Cardinaux, F., Kemp, T., Müller, K.-R., Curio, G., and Blankertz, B. (2016d). EEG-based usability assessment of 3D shutter glasses. *Journal of Neural Engineering*, 13(1):016003. doi: 10.1088/1741-2560/13/1/016003.

- Wessberg, J., Stambaugh, C. R., Kralik, J. D., Beck, P. D., Laubach, M., Chapin, J. K., Kim, J., Biggs, S. J., Srinivasan, M. A., and Nicolelis, M. A. L. (2000). Real-time prediction of hand trajectory by ensembles of cortical neurons in primates. *Nature*, 408(6810):361–365. doi: 10.1038/35042582.
- Wilkins, A. J. (1995). *Visual Stress*. Oxford University Press, Oxford, UK; New York, USA.
- Winkler, I. (2015). *On the extraction, classification and causal analysis of EEG signal components*. PhD thesis, Technische Universität Berlin. doi: 10.14279/depositonce-5055.
- Wolpaw, J. and Wolpaw, E. W., editors (2012). *Brain-Computer Interfaces: Principles and Practice*. Oxford University Press, Oxford, UK; New York, USA, 1st edition. doi: 10.1093/acprof:oso/9780195388855.001.0001.
- Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., and Vaughan, T. M. (2002). Brain-computer interfaces for communication and control. *Clinical Neurophysiology*, 113(6):767 – 791. doi: 10.1016/S1388-2457(02)00057-3.
- Wolpaw, J. R., McFarland, D. J., Neat, G. W., and Forneris, C. A. (1991). An EEG-based brain-computer interface for cursor control. *Electroencephalography and Clinical Neurophysiology*, 78(3):252–259. doi: 10.1016/0013-4694(91)90040-B.
- Woods, A. (2010). Understanding crosstalk in stereoscopic displays. Keynote presentation at Three-Dimensional Systems and Applications conference. http://cmst.curtin.edu.au/wp-content/uploads/sites/4/2016/05/2010-23_understanding_crosstalk_woods.pdf.
- Zander, T. O. and Kothe, C. (2011). Towards passive brain-computer interfaces: Applying brain-computer interface technology to human-machine systems in general. *Journal of Neural Engineering*, 8(2):025005. doi: 10.1088/1741-2560/8/2/025005.
- Zander, T. O., Lehne, M., Ihme, K., Jatzev, S., Correia, J., Kothe, C., Picht, B., and Nijboer, F. (2011). A dry EEG-system for scientific research and brain-computer interfaces. *Frontiers in Neuroscience*, 5(53):1–10. doi: 10.3389/fnins.2011.00053.
- Ziehe, A. and Müller, K.-R. (1998). TDSEP – an efficient algorithm for blind separation using time structure. In Niklasson, L., Bodén, M., and Ziemke, T., editors, *ICANN 98, Perspectives in Neural Computing*, pages 675–680. Springer, London, UK. doi: 10.1007/978-1-4471-1599-1_103.
- Ziehe, A., Müller, K.-R., Nolte, G., Mackert, B.-M., and Curio, G. (2000). Artifact reduction in magnetoneurography based on time-delayed second-order correlations. *IEEE Transactions on Biomedical Engineering*, 47(1):75–87. doi: 10.1109/10.817622.

The author wants to thank the École Polytechnique Fédérale de Lausanne for providing the well structured document template, which was used for this dissertation.