



Masters' Thesis

Catching the eye:

Investigating the functional neuroanatomy
of the visuospatial attention system with
fMRI and eyegaze recordings during natural
stimulation

submitted by

Adina Wagner (215486)

Master of Science in Clinical Neuroscience

adina.wagner@t-online.de

submitted to

J.-Prof. Dr. Michael Hanke

Prof. Dr. Stefan Pollmann

Psychoinformatics Lab

Otto-von-Guericke Universität Magdeburg

February 3rd, 2019

Statutory Declaration

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

Hiermit versichere ich, dass ich die vorliegende Arbeit selbständig und nur unter Benutzung der angegebenen Literatur- und Hilfsmittel angefertigt habe. Wörtlich übernommene Sätze und Satzteile aus anderen Druckwerken oder aus Internetpublikationen sind als Zitat belegt, andere Anlehnungen hinsichtlich Aussage und Umfang unter Angabe der Quelle kenntlich gemacht. Die Arbeit wurde in gleicher oder ähnlicher Form in keiner anderen Lehrveranstaltung als Leistungsnachweis eingereicht. Ich bin darüber unterrichtet, dass die Lehrenden angewiesen sind, schriftliche Arbeiten zu überprüfen, und dass ein Vergehen eine Meldung beim Prüfungsausschuss der Fakultät zur Folge hat, die im schlimmsten Fall zum Ausschluss aus der Universität führen kann.

Magdeburg, February 3rd, 2019

(Adina Wagner)

*The cure for boredom is curiosity. There is no cure
for curiosity.*

– Ellen Parr

Contents

Statutory Declaration	2
Table of Contents	4
List of Figures	6
List of Tables	7
List of Algorithms	8
List of Abbreviations	9
Acknowledgements	10
Abstract	1
1 Introduction	2
1.1 Visuospatial attention and the frontal eye fields	4
1.2 Movements of the eye	6
1.3 Insights on function and location from naturalistic stimulation	8
1.4 Utilizing eye movements in the study of attentional deployment	10
1.5 Investigating functional specificity with functional sensitivity profiles	11
1.5.1 Classification analysis	12
1.5.2 A "transposed" approach to the functional sensitivity of ROIs	14
2 Methods	16
2.1 Data basis	16
2.1.1 Stimulus material	16
2.1.2 Procedures	17
2.1.3 Software	17
2.1.4 Preprocessing	18
2.1.5 Extraction and processing of eye-tracking data	18
2.2 MultiMatch	19
2.2.1 multimatch in Python	19
2.2.2 multimatch usage in the studyforrest dataset	23
2.2.3 Deriving a measure of attentional modulation from multimatch	24
2.3 A novel approach for the analysis of ROI specificity	25
2.3.1 Method validation	26
2.4 TODO: FEF	27
3 Results	28
3.1 Eye movements and multimatch	28
3.2 Method validation	29

3.3	TODO FEF	31
4	TODO Discussion	32
4.1	TODO Drawbacks and shortcomings	33
4.2	TODO Future directions	34
4.3	TODO Conclusions	35
5	Things that are still shitty	36
	Bibliography	37
	Appendices	42

List of Figures

1.1	Dorsal and ventral attention networks	5
1.2	Regions involved in saccadic eye movements.	8
1.3	A linear decision boundary for a linear Gaussian Naive Bayes classifier .	14
2.1	Geometric representation of eye movements	20
2.2	Scanpath alignment as a shortest-path problem.	21
2.3	Similarity measures	21
3.1	Overview of extracted scanpaths' characteristics.	29
3.2	Functional specificity of ROIs in two validation analyses.	30

List of Tables

1.1	Movements of the eye	7
3.1	Characteristics of scanpath similarity across the movie.	28

List of Algorithms

1	multimatch	22
2	The studyforrest specific functions of multimatch	24

Nomenclature

EBA	Extrastriate Body Area
FEF	Frontal Eye Field
FFA	Fusiform Face Area
fMRI	functional Magnetic Resonance Imaging
GLM	General linear model
GNB	Gaussian Naive Bayes [classifier]
IFG	Inferior frontal gyrus
IPS	Intraparietal Sulcus
LOC	Lateral Occipital Complex
MEG	Magneto-encephalography
MF	Magnification Factor
MFG	Medial frontal gyrus
MSE	mean squared error
OLS	ordinary least-squares
PET	Positron Emission Tomography
PPA	Parahippocampal Place Area
SEF	Supplementary Eye Fields
SSE	explained sum of squares
SSR	residual sum of squares
SST	total sum of squares
STG	superior temporal gyrus
STS	superior temporal sulcus
TMS	Transcranial magnetic stimulation
TPJ	Tempoparietal Junction

Acknowledgements

I am thankful to a large number of people who supported and challenged me on the path of my master thesis completion and beyond. First and most important of all, I would like to thank my scientific supervisor J.-Prof. Dr. Michael Hanke of the Psychoinformatics Lab. He introduced me to the mesmerizing field of Psychoinformatics and has provided me with more opportunities I can recall to take root in it. A list of things that would not have been possible without him would span pages. I attribute much of my personal growth and scientific advancement of the past 1.5 years to him, and I am very grateful to continue pursuing my academic career as a PhD student under his supervision.

Secondly, I am grateful to Prof. Dr. Pollmann for agreeing to be the second assessor of my Masters Thesis, and for valuable discussions and insights with regard to the thesis. I cannot overstate the importance of Prof. Yaroslav Halchenko, whom I want to thank deeply for his agreement to supervise and mentor me on a four month research stay at Jim Haxbys Lab at Dartmouth College, NH, USA. His and Michaels passion for open science, open-source software, and scientific integrity and reproducibility will continue to be my source of inspiration and motivation throughout my scientific career.

I further am grateful to Prof. Jim Haxby for welcoming me in his lab in Dartmouth, as well as Prof. Ida Gobbini, Vassiki Chauhan, Kyle Meyer and Feilong Ma for their company and advice throughout my stay in the US.

This Masters Thesis would not have been possible without the methodological discussions, programming tutorials, and most importantly friendship of Alexander Waite, Dr. Emanuele Porcu, Benjamin Poldrack, Christian Häusler and Asim Dar. All members of the Psychoinformatics Lab have contributed to the most welcoming, fun, and empowering work environment I have ever been a part of, and despite my daily commute, this wonderful lab became a family for me.

All of my studies, in Germany and abroad, were mainly funded by the German Academic Foundation. The ideal and financial support since 2013 enabled me to become the person I am today. I am thankful to all of my referees, Dr. Ludwig, Dr. Trebesius, Dr. Köhne, and Dr. Chwalleck, for the insightful discussions, and my mentors, Prof. Enders and Prof. Fink, for their assistance and guidance, and to Dr. Julius and Prof. Zimmermann, whom I was honored to meet and work with several times.

Last but not least I am heartily grateful to my parents and friends for their patience, their support, their steady encouragement and constant moral guidance. Especially Gunnar Behrens, who was my anchor throughout seven busy years.

TODO: Abstract

[TODO]

[Introduce the topic]

[short statement of relevance - what is a gap in the literature]

[what is the aim of the thesis]

The analysis attempts to shed light on the functional specificity of one particular part of the visuospatial attention system, the frontal eye field [FEF]. For this aim, a multi-step procedure was employed. In a first step, a novel method to derive functional specificities of regions of interests was implemented and validated on data from a standard localization paradigm and a movie watching paradigm of the studyforrest dataset. In parallel, eye gaze data of participants were utilized both to derive a measure of visuospatial attention, and saccadic eye movements. [We're still missing the main approach here...] [what was the main result]

[if there's any space left, conclusions.]

Keywords: *fMRI, eye-tracking, naturalistic stimulation, methods, localization*

1 Introduction

In complex natural environments, attention is necessary to handle an excess of information despite limited neural computing power (Carrasco 2011). It serves as a filter to select potential behaviorally or cognitively relevant cues from the continuous stream of information, and as a gatekeeper to discard irrelevant stimuli in order to prevent sensory overload (Bellebaum, Thoma, and Daum 2012). *Visual attention* provides this filtering function for visual perception: Relevant visual stimuli are prioritized and hence attended while less important aspects of a scene are neglected. In this way, visual attention is crucial for the selection of cognitive and behavioral strategies for interacting with the environment (Siegelbaum and Hudspeth 2000).

One integral functional region involved in both the selection and attendance of behaviorally relevant aspects of the visual environment is the *frontal eye fields* (FEF). They serve core functions in the movement of the eyes (Krauzlis 2014), and within the dorsal endogenous attention system (Corbetta and Shulman 2002; Corbetta, Patel, and Shulman 2008). While well localized and understood in animal models via electric stimulation (e.g. Bruce et al. 1985), non-invasive research into the location and function of human FEF under natural conditions is more challenging. Current literature displays a remarkable diversity of locations between subjects and methods (see 1.3, Vernet et al. 2014), and highly controlled experimental conditions lacking the ecological validity of natural stimulation may fail to evoke and study the attentional modulation of the FEF.

This thesis aims to enrich current literature on the functional neuroanatomy and location of the FEF by means of methodological advancements in two different aspects: For one, it attempts to utilize information derived from eye-tracking data for a measure visuospatial attention deployment. Secondly, it introduces a novel approach for determining the specificity of functional regions of interest. Lastly, it attempts to combine both aspects to gain insights into the location and attentional modulation of the FEF during naturalistic stimulation with the Hollywood movie *Forrest Gump*.

This thesis is structured as follows. Chapter 1 gives an overview of visuospatial attention and its underlying neuroanatomical foundation with an additional focus on the frontal eye fields (section 1.1), their relation to the movements of the eye (section 1.2), and the utilization of eye movements in the study of attentional deployment (section 1.4). Furthermore, it outlines current localization paradigms with their potential shortcomings, and describes a novel method to derive functional information from an ROI localization (section 1.5). Chapter 2 introduces the data basis. It elaborates on the chosen method for scanpath comparison, *multimatch*, and the developed ROI specificity analysis and its validation. Chapter 3 presents the results of the analyses, and chapter

4 discusses their implications and limitations. **TODO: Extend this introduction to include more about the motivation and aims – this is too short.**

1.1 Visuospatial attention and the frontal eye fields

The exogenous (greek *exo* = outside, *genein* = to produce) visual attention, also referred to as involuntary or bottom-up controlled, visual attention, is driven by the context-dependent saliency of external stimuli (Itti and Koch 2001). If the physical visual properties like color, luminance, motion or contrasts of a stimulus are conspicuous enough given the context they are in, they are able to capture the visual attention even if the observer doesn't intent to orient attention to it (Chica, Bartolomeo, and Lupiáñez 2013). According to Corbetta and Shulman (2002) attention framework, a ventral attention network consisting of the tempoparietal junction (TPJ)/superior temporal sulcus and gyrus (STS, STG), the medial/inferior frontal gyri (MFG/IFG), and the ventral part of the supramarginal gyrus (SMG) processes attentional selection based on such exogenous cues (Figure 1.1, orange areas), if they are behaviorally relevant (Downar et al. 2000). While there is some differential evidence for the TPJ (see Vossel, Geng, and Fink (2014), for a short overview) a majority of studies showed a lateralization of the ventral system to the right hemisphere (Corbetta and Shulman 2002; Fox et al. 2006; Corbetta, Patel, and Shulman 2008).

The endogenous (greek *endo* = within) visual attention, also termed voluntary or top-down controlled visual attention, on the other hand is controlled by selection criteria that depend on cognitive factors such as requirements of the current task, prior experiences or expectations (Itti and Koch 2001). This goal-directed selection is controlled by a dorsal frontoparietal network including the intraparietal sulcus (IPS) and superior parietal lobule (SPL), and the dorsal frontal cortex along the precentral sulcus, near or at the frontal eye field (FEF) and the supplementary eye field of each hemisphere (SEF; figure 1.1, blue areas; Corbetta and Shulman 2002; Corbetta, Patel, and Shulman 2008). Endogenous attention is usually drawn to only weakly salient stimuli that would not evoke exogenous attention but are relevant for current cognitive aims. As such, whereas exogenous attention is a rapid, automatic orientating response, endogenous attention is a voluntary and controlled process of information monitoring (Carrasco 2011).

A core part of the dorsal attention system is the FEF. In humans, the FEF is located bilaterally in the rostral bank of a portion of the precentral sulcus at the caudal end of the middle frontal gyrus. Early insights about its function stem from stimulation studies with primates and were related to the FEF functions in the generation of eye-movements, in particular saccades: Low-intensity electrical stimulation of the FEF in the bank of the arcuate sulcus in monkeys elicits saccadic eye movements, directed contralaterally to the stimulated hemisphere (Tehovnik et al. 2000). Work by Bruce et al. (1985) further found that saccades elicited by stimulation at a particular region have a particular direction and amplitude that is independent of the orientation of the eyes, thus finding the first evidence of a subdivision within the FEF. Their work revealed a gradient of saccadic amplitudes from lateral to medial sites of the FEF: The more medially the stimulation of the primate arcuate sulcus, the larger the saccadic amplitude (Bruce et al. 1985). Hence, the ventrolateral portion of the FEF is gener-

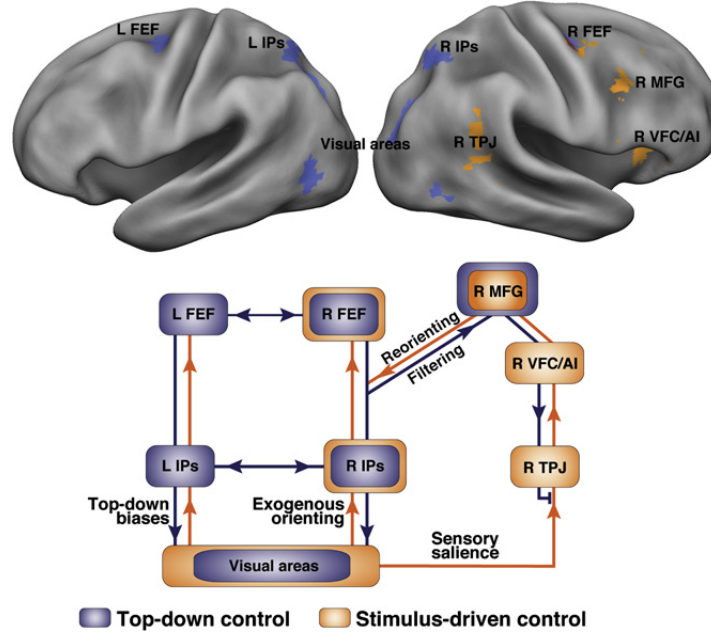


Figure 1.1: Definition of dorsal and ventral networks from activation data and putative interactions. Taken from Corbetta, Patel, and Shulman (2008).

ating shorter saccades, while the mediodorsal portion is generating longer saccades. Extensive connectivity between the FEF and the visual cortex or other areas related to vision such as the dlPFC provide evidence for a substantial role of the FEF in vision (Stanton, Bruce, and Goldberg 1995). Even the aforementioned partialization of the FEF is propagated within connections with other visual areas caudal to the central sulcus. The more ventrolateral portion of the FEF, responsible for generating shorter saccades, is interconnected with retinotopically organized areas that represent central vision (i.e. within 10° of the fovea) in the inferotemporal cortex. In contrast, the mediodorsal FEF, responsible for generating longer saccades, is interconnected with areas with a representation of the peripheral visual field (Stanton, Bruce, and Goldberg 1995, Schall et al. 1995). Stimulation studies in primates and high resolution fMRI experiments in humans further revealed that the FEF also contains zones that are involved in the control of other types of eye movements (see Krauzlis (2014) for an overview): Stimulation of the frontal pursuitzone in the fundus elicits pursuit eye movements directed ipsilateral to the stimulated hemisphere (Blanke and Seeck 2003). The intracortical stimulation of several subareas within the FEF, among others a region within the pre-arcuate cortex in rhesus monkeys, immediately rostral to the saccade related anterior bank of the arcuate sulcus, triggers vergence movements (changes of the depth of the gaze), and shows an involvement in accommodation (Crosby et al., 1952, cited by Vernet et al. (2014)). The FEF is further active in the disengagement from fixations prior to a new saccade (Goodwin 2007; Tehovnik et al. 2000). It therefore plays a crucial motor function in most types of eye movements, most prominently saccades, and also pursuit, fixation, and vergence eye movements.

However, as its involvement in the dorsal attention network suggests, the FEF is fur-

ther involved in a number of aspects of higher cognition, such as attention. The most compelling evidence stems from lesion studies: unilateral damage to the FEF results in contralateral neglect (Crowne, Yeo, and Russell 1981). In healthy humans, Mugleton et al. (2003) found that repetitive transcranial magnetic stimulation (rTMS) at 10Hz for 500ms over the right FEF impairs only subtypes of visual search involving endogenous attention, namely where a visual target is neither salient nor predictable. Further studies of both attentional and motor processes of the FEF found that the FEFs functions in motor control and attention can be partially dissociated: activation in the FEF reflects deployment of attention even in the absence of accompanying eye movements to the region of interest, so called *covert attention* (Vossel, Geng, and Fink 2014), and medical inactivation of the FEF in monkeys disrupts covert visual search tasks (Monosov and Thompson 2009). Nevertheless, attentional processes and motor functions of the FEF seem closely related. Thompson and Bichot (2005) studied the FEFs role in saccade target selection in primates performing a visual search task. They found evidence for the existence of a heatmap-like representation of saliency encoded in selective activation that is related to the overall behavioral relevance of visual stimuli. This representation, the authors argue, arising from an integration of both top-down and bottom-up influences, then determines targets for foveation in the visual field in a 'the-winner-takes-it-all' fashion.

The frontal eye fields are therefore an interesting structure occupying important roles in aspects of higher cognition, in particular visuospatial attention, eye movements, and their interaction: Following Vernet et al. (2014) reasoning and the short overview given in this section, studies of the frontal eye field in conjunction with the cognitive context of the viewer could shed light on how it modulates the activation of the FEF and its involvement in the various types of eye movements, and gain insight on its role as a site of conversion for the selection of visual targets for eye movements and visuospatial attention (Corbetta and Shulman 2002; Kowler 2011).

TODO: THIS NEEDS A BETTER OUTRO.

1.2 Movements of the eye

The close interaction between attention and eye movements is further eminent in the underlying neuroanatomical principles of human vision. A central feature of the human eye is the fovea centralis, a specialised region about 1.5mm in diameter in the center of the retina (Drenckhahn 2004). It serves only the central 1° of the visual field, but due to possessing the highest amount of cones of the retina, and an asymmetric distribution of ganglion cell density across the retina that advantages foveal information, it provides the greatest visual acuity (Perry and Cowey 1986). This asymmetry is propagated in the cortical representation of visual inputs from the retina. The magnification factor (MF), the linear extent of cortex devoted to each linear degree on the retina, increases monotonically from periphal to foveal vision (Daniel and Whitteridge 1961). As a result, in almost all visual brain areas, both cortically and subcortically, the fovea has the greatest representation, and full visual acuity can therefore only occur at the fovea.

Due to this constraint, humans need the ability to, first, align the fovea rapidly to an object of interest in a visual scene and, second, keep the fovea aligned to it for a sufficient amount of time in order to maximize the efficiency of foveal vision and perceive objects of interest in greatest detail. An exploration of a visual scene is hence performed in a stepwise manner and requires the described sequence of eye movements to sequentially explore all areas of interest. To accomplish this, the human eye is capable of a number of different eye movements. In its most simple form, the sequence consists of rapid *saccadic* eye-movements, redirecting the fovea from one object of interest to the next, followed by phases of *fixation* that keep the fovea aligned for visual analysis. As saccades reach speeds of up to 500 °/sec, small, corrective post-saccadic oscillations, so called *glissades*, correct minor deviations between saccade target and actual saccade endpoint. Additionally, *smooth pursuits* enable a viewer to track moving objects, while *vergence* movements are necessary to perceive different depths (Holmqvist et al. 2011). The following table 1.1 provides an overview of these movements and their characteristics. Note, however, that most work on eye movements was conducted on static stimuli, and eye movements on moving images could possess differential characteristics. Dorr et al. (2010) for example investigated fixation characteristics differentially for varying stimuli and found that Hollywood movies elicit longer fixations (mean: 354ms, median: 253ms) than static images.

Type	Duration(ms)	Amplitude (degrees)	Velocity (°/sec)
Fixation	200-300	-	-
Saccade	30-80	4-20	30-500
Glissade	10-40	0.5-2	20-140
Smooth pursuit	-	-	10-30

Table 1.1: An overview of major types of eye movements and their characteristics, taken from Holmqvist et al. (2011).

With regard to the scope of this thesis, and as eminent from the fundamental principles of foveal vision, saccades are the type of eye movement of particular relevance to the study of visuospatial attention (Kowler 2011). As mentioned in the previous section, the FEF, core part of the dorsal attention network, is playing a key role in it. Saccadic eye movements are controlled by a number of highly interconnected brain regions, both cortical and subcortical (Munoz 2002, see figure 1.2), but in the cortex, the FEF are seen as the principal region involved in their oculomotor control (Leichnetz and Goldberg 1988). Thus, in regard to the study of the functional neuroanatomy of the visuopatial attention system by means of eye movement, saccades appear as a plausible choice of target movements.

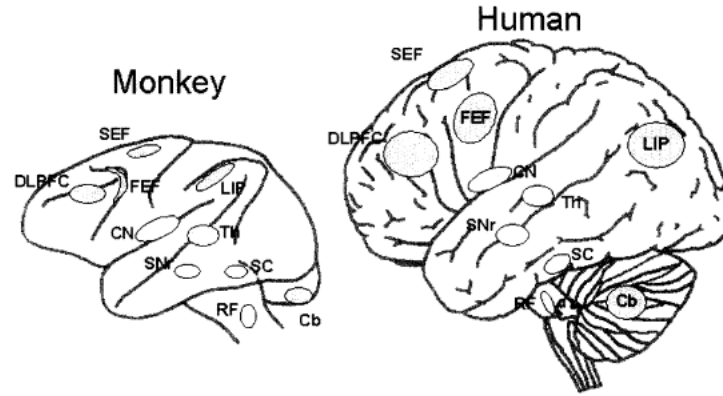


Figure 1.2: Brain areas involved in the control of saccadic eye movements in monkey (left) and human (right). *Cb*: Cerebellum, *CN*: Caudate nucleus, *DLPFC*: dorsolateral pre-frontal cortex, *FEF*: Frontal eye field, *LIP*: Lateral intraparietal area; *RF*: Reticular formation, *SC*: superior colliculus, *SEF*: supplementary eye field, *SNr*: substantia nigra pars reticularis, *Th*: Thalamus. (Taken from Munoz (2002)) **Todo:** this might be superfluous, kick out if in need of space

1.3 Insights on function and location from naturalistic stimulation

Most research paradigms in neuroscience employ highly controlled stimulation protocols to precisely pinpoint and disentangle the function or location of the structure in question (Malinen, Hlushchuk, and Hari 2007). This approach however can suffer from a number of shortcomings in the study of attention or in the localization and study of functional regions of interest. The following section outlines a number of these shortcomings and argues for the use of naturalistic stimulation for additional insights instead.

Visuospatial attention is usually studied by means of highly controlled experimental tasks, such as Posners' cueing paradigm (Posner, Snyder, and Davidson 1980) or simple visual search tasks. While well controllable, these conventional laboratory experiments suffer from a lack of ecological validity and fail to evoke realistic competing demands for visuospatial attention. As Hasson et al. (2004) noted, controlled experimental settings bear little resemblance to natural viewing for at least four reasons: 1) Lack of complex visual scenes (i.e. presentation of visual stimuli in isolation), 2) Lack of (complex) movement, 3) Lack of unconstrained eye movement, and 4) Lack of interactions between vision and additional modalities, context and emotional valence. Naturalistic stimulation such as watching movie clips, in contrast, does not suffer from these shortcomings. Therefore, while static images have been an object of interest in the analysis of gaze distribution and attention from the early works of Yarbus (1967) onwards, during the past decade, movies or videos became a promising stimulus choice: They contain interesting, multisensory stimuli and are more representative of natural vision arrays than static pictures. As movie content constantly changes and includes both

salient elements such as motion, as well as top-down components from the attempts to comprehend and interpret the content (Ross and Kowler 2013), movies represent an ideal possibility to study the dynamic and complex interplay of attentional processes ecologically valid (e.g. in Hasson et al. (2004); Carmi and Itti 2006; Tseng et al. 2009; Dorr et al. 2010).

Localization of ROIs is most commonly achieved by simple contrasts in a general linear model (GLM) between experimental conditions that are known to activate functional regions. To localize the fusiform face area (FFA) for example, a standard paradigm consists of the contrast in activation during block stimulation with images of faces, and images from a different category such as houses (Sengupta et al. 2016, Fox, Iaria, and Barton 2009). Localization of areas involved in motor tasks such as the FEF would most reliably be conducted via micro-stimulation, which is rarely possible in human subjects. Localization through non-invasive methods is more challenging, and usually employs simple behavioral paradigms with either general oculomotor tasks (Paus 1996), or more specific oculomotor tasks such as instructed fixations, pro- and antisaccades (Connolly et al. 2002), or tracking of horizontal step stimuli (Alkan, Biswal, and Alvarez 2011). However, while capable of localizing areas undoubtedly involved in the movement of the eyes, the displayed movements are not performed under unconstrained conditions. For regions involved in higher cognition such as the FEF it remains unknown whether uncontrolled eye movements under naturalistic viewing behavior can be capable of providing similar, if not even additional information when taking cognitive context such as attention deployment into account. Previous localization studies employing positron emission tomography (PET) (e.g. Paus (1996), Kawashima et al. 1998), magneto-encephalography (MEG) (e.g. Ioannides et al. 2004), fMRI (e.g. Petit and Haxby 1999; Connolly et al. 2002), and transcranial magnetic stimulation (TMS) (see Vernet et al. (2014), for an overview) were able to shed light on the possible location of the FEF in humans. However, an overall view of this literature also reveals the large variability of reported localizations between studies (see e.g. Paus (1996); Vernet et al. (2014), for overviews), and there is no consensus on whether these discrepancies arise from methodological differences in the choice of imaging technique or behavioral paradigm, or inter-individual differences between participants. A more naturalistic paradigm might be able to provide additional answers to this question.

To address the blank spaces left by paradigms with highly controlled, simplistic stimuli or tasks, the present thesis aims to use a more ecologically valid stimulation paradigm by using fMRI and eye-tracking data obtained during movie watching. Due to its ready availability, data richness, and employment of naturalistic stimuli in form of the Hollywood movie "Forrest Gump", the studyforrest datasets were selected as the data basis of this thesis. **TODO: this outro just reads super stiffly**

1.4 Utilizing eye movements in the study of attentional deployment

In order to include the aspect of visuospatial attention, a quantitative measure of attentional deployment needs to be derived. One possibility to do this is by using eye gaze information in the form of *scanpaths*. The term scanpath refers to the trace of eye-movements in space and time (Holmqvist et al. 2011). In its most simple form, it is formed by a succession of fixations and saccades that define the particular sequence in which the eyes explore a visual scene (Anderson et al. 2015). As opposed to other measures that summarize eye gaze, such as heatmaps, the order of eye-movements is relevant - a different order of elements in the representational sequence of eye-movements constitutes a different scanpath.

The analysis of scanpaths has been used for decades to gain insights into the viewers' mental processes, especially those concerning visuo-spatial attention. Yarbus (1967) wrote:

„Eye-movements reflect the human thought processes; so the observer's thoughts may be followed to some extent from records of eye-movements (the thought accompanying the examination of the particular object). It is easy to determine from these records which elements attract the observer's eye (and, consequently, his thought), in what order, and how often.“

Following this reasoning, the comparison of scanpaths of different subjects constitutes a measure of similarity in the different subjects' attentional processes and adds a useful dimension to the traditional analysis of eye-tracking data.

In recent years, many approaches for scanpath comparisons were developed and implemented in various software solutions. Among them are methods based on (semantic) areas of interest (AOIs) such as the *Levenshtein distance* (Levenshtein 1966), or an improved generalization of it in *ScanMatch* (Cristino et al. 2010), methods based on attention maps such as *AMAP* (Ouerhani et al. 2004), methods employing machine learning algorithms such as *SMAC with HMM* (Coutrot, Hsiao, and Chan 2018), or methods based on vector-geometry such as *MultiMatch* (Jarodzka, Holmqvist, and Nyström 2010). The latter method is a multidimensional approach of scanpath comparison on five different measures similarity, **shape**, **length**, **position**, **direction**, and **fixa-tion durations**. According to Jarodzka, Holmqvist, and Nyström (2010), it has been developed specifically to overcome known shortcomings of many previous methods that limit their informative value. These shortcomings amount to a loss of information due to the use of coarse aggregate measures of eye movement, susceptibility of the results of a comparison to influential data points due to arbitrary classifications, or both. The main reason it overcomes these deficiency is that it does not rely on AOIs, but the precise locations in pixels in two-dimensional space, which leads to a finer level of detail in scanpath comparison.

A number of studies used the method in evaluations or applications of scanpath comparisons. In evaluations with simulated and actual eye-tracking data, the method was

found to be robust against spatial noise, sensitive to position, order and fixation duration, and outperformed the ScanMatch method in AOI border cases (Dewhurst et al. 2012). In a comprehensive test of eleven common scanpath comparison methods on static real-life photographs, MultiMatch was found to be robust in inter- and intra-subject comparisons of scanpaths (Anderson et al. 2015). As it was also used to study mental activity involved in perceiving visual input by others (French, Gladly, and Thibaut 2017) already, the MultiMatch method was chosen to be employed in the current thesis to derive a measure for attentional processes of participants during movie watching.

For its computation, this attentional measure utilizes the scanpath information of all participants to obtain a common measure of attentional modulation of the stimulus in the sample. It follows a reasoning conceived by **Baumgartner et al. (unpublished manuscript)**: The more similar all participants scanpaths are for any given scene of a movie stimulus, the more exogenous control is exercised by this movie segment.

TODO: this might need additional details/motivation/examples

An overview of the method is available in section 2.2.

1.5 Investigating functional specificity with functional sensitivity profiles

In addition to using scanpath comparisons for a measure of visual attention, a novel method to investigate the functional specificity of regions of interest is developed and implemented in this thesis. As outlined in section 1.3, a general approach of localization consists of contrasting experimental conditions known to activate specific functional regions in a GLM, for example the contrast between activation during static images of houses and static images of faces to localize the FFA (Fox, Iaria, and Barton 2009), or the use of instructed horizontal saccades versus fixations to localize the FEF (Connolly et al. 2002). This approach suffers from a number of drawbacks:

First, as outlined as well in 1.3, the employed artificial stimulation is not ecologically valid, which decreases the sensitivity of functional localizers and thus their applicability across studies. Secondly, a reliance on narrow stimulus categories and simple contrasts omits insights into further stimulus aspects of potential relevance - especially since an exact contrast between conditions accounting for various ROI-specific factors is not known ahead of time, and thus cannot be incorporated into the localizer GLM contrast. Moreover, it is challenging to differentiate functionally related but different ROIs with a simplistic experimental design. Therefore, localization often further requires to take spatial information for functionally close regions (such as the occipital face area (OFA) and the FFA) into account. Lastly, changes in statistical thresholds can alter localization results substantially (Fox, Iaria, and Barton 2009) and introduce an additional source of *researchers degrees of freedom* into an analysis. Obtaining additional insights in regard to functionality of regions of interest might be beneficial to reduce reliance on spatial information, subjective thresholding, and may additionally

give rise to new insights into the functional neuroanatomy of ROIs.

The approach introduced in this thesis is designed to estimate a maximally discriminative contrast between the functional signatures, i.e. the activation time courses, of any two ROIs in question, and to use the available experimental design - regardless of complexity - to provide a functional description for it. The results of this analysis should give differential insights into the functional specificity of ROIs beyond simple contrasts of stimulus conditions.

The method relies on sensitivity profiles derived during a classification analysis. A short introduction into the basic principles of this type of machine learning tool and the chosen algorithm for this analysis will be introduced in the next subsection, followed by an elaboration on how the derived sensitivities are further utilized. In order to validate the method, its applicability and generalization capabilities were first tested on a general localization paradigm, and later movie-data, for already localized ROIs, which will be elaborated on in section 2.3.1.

1.5.1 Classification analysis

The term *classification analysis* is most often used in the context of *machine learning*. Machine learning refers to the study and construction of software that can learn from data without being explicitly programmed (Zeigermann and Nguyen 2018). Especially in the data-rich environment of neuroimaging, machine learning algorithms and techniques have gained immensely in popularity and significance as for their ability to detect hidden patterns and trends in the data (Vogt 2018). One particular type of tool from machine learning is a so called *classifier*, which is used for a classification analysis. In such an analysis, a classifier is trained on a subset of data, the *training set*, and is then used for predictions on a previously unseen subset of data, the *test set* (Zeigermann and Nguyen 2018).

More formally, a classifier is a function created from training data to predict the value of a class attribute $C \in \{1, \dots, r\}$ (the *label*), given the predictive *features* $X = (X_1, \dots, X_d)$ with values in $S = \{S_1 \times \dots \times S_d\}$ and the sample $x = (x_1, \dots, x_d)$. Suppose (X, C) is a random vector with a joint feature-label probability distribution $p(x, c)$. A classifier ψ is the function that maps X onto C :

$$\psi : \{S_1, \times \dots, \times S_d\} \rightarrow \{1, \dots, r\} \quad (1.1)$$

$$x \mapsto c. \quad (1.2)$$

The overall aim of the classification analysis in this thesis was not the classification in general, but to derive a sensitivity measure in the form of feature weights from a simultaneous sensitivity analysis. There are a number of different classification algorithms available to build classification models, and based on prior work of **Nastase et al. (2016)**, a *linear Gaussian Naive Bayes* classifier (GNB) was further explored due to its high classification accuracy and easy interpretability of feature weights in case of linearity.

The algorithm belongs to the family of Naive Bayes algorithms. These algorithms use

a probabilistic approach to classification, utilizing the Bayes theorem to capture the relationships within input data and output label. Given some samples x_1, \dots, x_n with a number of characteristics on different features, and labels c_1, \dots, c_r , naive Bayes classifiers learn a model of the joint probability, $p(x, c)$ during training. During testing, they make predictions by using Bayes theorem to calculate the probability of a label, given the feature characteristics of a sample, $p(c|x)$. Two fundamental assumptions exist for these algorithms:

- *Conditional independence*: Each feature x_i is independent of every other feature x_j for $j \neq i$ given the outcome c_r .
- *Equal contribution of features*: Each feature is given the same weight for the outcome prediction.

A linear Gaussian Naive Bayes classifier further assumes the data to be distributed according to a Gaussian distribution, and it assumes the feature variance to be independent of class (i.e. $\sigma_{i,0}^2 = \sigma_{i,1}^2$). Given the latter assumption holds, the decision boundary between any two classes 1 and 0,

$$\prod_{i=1}^d P(X_i|C=0)P(C=0) = \prod_{i=1}^d P(X_i|C=1)P(C=1), \quad (1.3)$$

is linear (for the mathematical proof, see e.g. Mitchell (1997)), chapter 3). The feature weights of this classifier are defined as

$$w_i = \sum_i \frac{\mu_{i,0} - \mu_{i,1}}{\sigma_i^2}. \quad (1.4)$$

TODO: Draw a picture with a linear decision boundary between two classes with two features, plus the Gaussian distributions and equal variances, just for clarity. this part is not comprehensible at all for anyone not in the topic...there are means to get this done in Latex, packages are imported already, but might be time consuming...or plot it with pymvpa... for now the gimp one suffices. Steal idea for visualization from:

<https://www.youtube.com/watch?v=TcAQKPgymLE>, just make it linear.

Figure 1.3 illustrates the case of two features and two classes with their respective distribution. The feature weight of the feature A can be derived as the difference in means given the class, $\mu_{A,1} - \mu_{A,2}$, divided by their common variance.

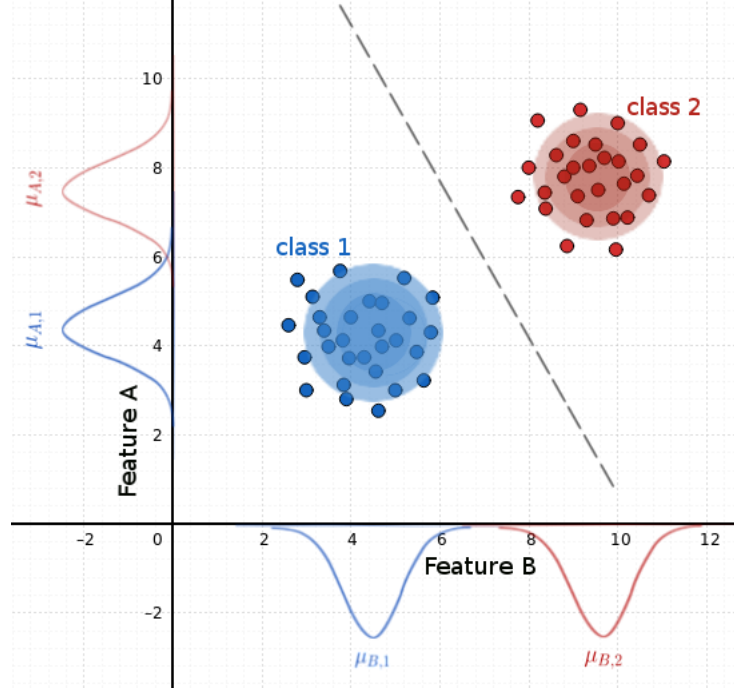


Figure 1.3: Exemplary illustration of a linear decision boundary in the case of equal variances. The distribution of each features data given the class follows a Gaussian distribution with mean μ_i and variance σ^2 .

1.5.2 A "transposed" approach to the functional sensitivity of ROIs

The method developed here utilizes the sensitivities obtained during a classification analysis and thus extends work by **Nastase et al. (2016)**: Following a classification of voxels (samples) to ROIs (labels) based on the observed data in the ROIs from other participants, a subsequent analysis of the classifiers sensitivity profile obtained from a decision between ROI pairs is used to shed light onto their functional distinctiveness. As opposed to the common approaches at classification with subsequent sensitivity analysis in neuroimaging (e.g. as used in Poldrack, Halchenko, and Hanson (2009)), the datasets are transposed: The features in this classification problem are time points, while voxels constitute samples.

In a first analysis step, a sensitivity profile of two ROIs is obtained via classification with the linear GNB. As the dataset is transposed to have time points as features, this sensitivity profile is a time course of sensitivities: each weight corresponds to the difference in mean activation of all voxels given the label (ROI) at one time point, normalized by their common variance. In an exemplary decision between ROIs A and B, a negative weight corresponds to an overall higher activation in B at this particular point in time. In a second step, this time course is entered as a criterion in a GLM with details about the experimental design at different time points as regressors. The resulting coefficients indicate the importance for each regressor (i.e. stimulus characteristic) for the distinction between ROIs. Section 2.3.1 will give a comprehensive

example of this approach.

TODO: [HOW CAN WE USE THIS FOR FEF]

2 Methods

The following method section attempts to give a concise description of this thesis’ data basis, fMRI and eye-tracking, and the employed and developed methods, MultiMatch and ROI specificity analysis. **[TODO: ... and whats presumably the main part of this thesis, the FEF thingy]**

2.1 Data basis

Data stems from the 2016 released extension of the studyforrest dataset¹ (Hanke et al. 2016; Sengupta et al. 2016). In this extension, $N = 15$ right-handed participants (age range 21 – 39 years, mean age 29.4 years, six female, normal or corrected-to-normal vision), who had previously participated in the studyforrest project, watched the audiovisual movie “Forrest Gump” (R. Zemeckis, Paramount Pictures, 1994) during simultaneous fMRI and eye-tracking recording. They further underwent a traditional localizer paradigm for higher visual areas: The fusiform face area (FFA), the occipital face area (OFA), the extrastriate body area (EBA), the hippocampal place area (PPA), the lateral occipital complex (LOC) and early visual cortex (Sengupta et al. 2016).

2.1.1 Stimulus material

Stimulus material for the audiovisual movie was the German dubbed version of Forrest Gump, overlaid with a German audio-description (**Koop, Michalski, Beckmann, Meinhardt & Benecke**, produced by Bayrischer Rundfunk, 2009), originally broadcast as an additional audio track for visually impaired recipients on Swiss public television. The additional audio track consists of a male narrator describing the visual content of a scene between dialogs, off-screen-speech or any other relevant audio-content of the movie (Hanke et al. 2014). The video track for the movie stimulus was re-encoded from Blu-ray into H.264 video (1280 x 720 at 25 frames per second (fps)). In accordance to the procedure in an earlier phase of the studyforrest project, the movie was shortened by removing a few scenes less relevant for the major plot to keep the fMRI recording session under two hours. The shortened movie was then segmented into eight segments of roughly 15 minutes of length (for an overview on segment duration, final stimulus content and detailed procedures see Hanke et al. (2014)). Stimulus material for the localizer paradigm consisted of 24 unique gray-scale images for each of six different categories (faces, bodies, houses, scenes, everyday objects, and scrambled images)

¹all data and code are publicly available at <https://github.com/psychoinformatics-de/studyforrest-data-phase2>.

at a resolution of 400 x 400px, matched in luminance, and displayed at 10 x 10° of visual angle (Sengupta et al. 2016).

2.1.2 Procedures

Functional MRI data acquisition for the audio visual movie was undertaken in two consecutive recording sessions on the same day, with a break of flexible duration. Within each session, four movie segments were presented in chronological order (Hanke et al. 2016). Visual stimuli were projected on to a screen inside the bore of the magnet using an LCD projector, and presented to the subjects through a front-reflective mirror on top of the head coil at a viewing distance of 63cm. The screen dimensions were 26.5cm x 21.2cm (corresponding to 1280 x 1024 px) at a resolution of 720p at full width, with a 60Hz video refresh rate (Sengupta et al. 2016). Eye-tracking was performed with an Eyelink 1000 (software version 4.594) using monocular corneal reflection and pupil tracking with a temporal resolution of eye gaze recordings of 1000Hz. The camera was mounted at an approximate distance of 100cm to the left eye of subjects, which was illuminated by an infrared light source (Hanke et al. 2016). Movie presentation and eye-tracking were synchronized by starting the eye gaze recording as soon as the stimulus computer received the first fMRI trigger signal. Timings of subsequent trigger pulses and onsets of every movie frame were logged. Using a 3 Tesla Philips Achieve dStream MRI scanner with a 32 channel head coil, T2*-weighted echo-planar images (gradient-echo, TR = 2s, echo time = 30ms, flip angle = 90) were acquired during movie watching. For the eight segments, 451, 441, 438, 488, 462, 439, 542, and 338 volumes were acquired, respectively (Hanke et al. 2016).

In the localizer paradigm, participants viewed the different object categories in a total of four block-design runs, with two 16s blocks per stimulus category. The order of the individual images per category differed across runs and participants, but the sequence of category blocks was the same for every run and participant. Localizers for higher visual areas (FFA, OFA, PPA, LOC, EBA, and early visual cortex) were obtained by the original authors of the study forrest extension by means of a two-level GLM analysis on contrasts of activation between object categories (for details, see Sengupta et al. 2016).

2.1.3 Software

Data preprocessing and analysis were conducted with a multitude of Python-based software packages. Preprocessing of data relied primarily on custom made workflows written with nipy (Gorgolewski et al. 2017), and utilizing FSL (Jenkinson et al. 2012). Much of the data analysis was done using pyMVPA (Hanke et al. 2009). Datalad, a version control tool for scientific data (<https://zenodo.org/record/1470735>), served as a backbone for all own work presented in this thesis. If applicable and unless otherwise stated, the alpha level for statistical significance was set to $\alpha = 0.05$.

2.1.4 Preprocessing

Eye-tracking data were normalized such that all gaze coordinates are in native movie frame pixels, with the top-left corner of the movie frame located at (0, 0) and the lower-right corner located at (1280, 546) (Hanke et al. 2016). The amount of unusable data, primarily due to signal loss due to eye blinks, ranged from less than 1 to 15% for 13 of the 15 in-scanner subjects (the other two subjects' data contained 85 and 36% of data loss, respectively). In-scanner acquisition had an approximate spatial uncertainty of 40px according to the calibration procedure.

The raw eye-tracking data was classified into different eye movements. For this, based on an adaptive, velocity-based algorithm proposed by Nyström and Holmqvist (2010), **Dar et al. (in preparation)** implemented a data-driven algorithm for robust eye movement detection for natural viewing (REMoDNaV) in Python². All results of this algorithm are publicly available at [**TODO**: URL here]. The algorithm categorizes the raw data into the eye movement categories saccades, fixations, post-saccadic oscillations (glissades), and disregards any unclassifiable data (such as blinks). The eye events are reported together with their start- and end coordinates, their onsets and duration in seconds, their velocity and the average pupil size.

Motion-corrected fMRI data in a study-specific group-template were obtained from the respective Github repositories via datalad [**CITE? Zenodo thing is ugly**]. In a first step, both datasets, localizer and movie-data, were preprocessed identically using a custom made nipype workflow, consisting of whole-brain masking, high-pass filtering and smoothing with a Gaussian kernel of 4mm FWHM³.

TODO: this might need to become a tiny bit more detailed/informative

2.1.5 Extraction and processing of eye-tracking data

For the extraction of saccadic eye events, the results were filtered to include only saccades. For the computation of event files following the BIDS standard (Gorgolewski et al. 2016), the Cartesian coordinates of the saccades were transformed into Polar coordinates (angle and length, a graphic explanation of this procedure can be found in 2.1, B). Based on their angle, saccades were classed into event types describing eight different orientations in the visual field. Saccade onsets and durations were extracted from the REMoDNaV output. The resulting event file per run hence contained the onset, duration, amplitude (length) and direction of saccades.

TODO: and now what did we do with it? TODO: At one point this might just be integrated into another method section

²The sourcecode can be found at github.com/psychoinformatics-de/remodnav.

³All code for preprocessing and analysis can be found at <https://github.com/AdinaWagner/localizer>.

2.2 MultiMatch

The MultiMatch method (Jarodzka, Holmqvist, and Nyström 2010) was selected as a promising approach of scanpath comparison. The original implementation, however, only existed as a Matlab toolbox shared upon request by the corresponding author of the respective publication (Dewhurst et al. 2012). This reliance on closed-source software and barriers in retrieval of the software⁴ hinder the wider usage and potential improvement of the method. A publicly available, open-source implementation would instead facilitate its usage in scientific works. Therefore, the method was ported from Matlab into Python and transformed into the pip-installable module *multimatch*⁵.

2.2.1 multimatch in Python

The following section gives a brief overview of the general MultiMatch method. An overview in pseudocode is outlined in algorithm 1.

The method takes two $n \times 3$ fixation vectors of two scanpaths with x-coordinates in pixel, y-coordinates in pixel, and duration of fixations as columns as its input. Based on the coordinates and durations of fixations, the scanpaths are represented as geometric vector sequences as shown in figure 2.1:

⁴While acquisition of the toolbox via e-mail was prompt and complication-free, later correspondence with several of the authors was an odyssey of expired email addresses, changed affiliations and digital communication allegedly lost in the void of the Internet.

⁵The module can be installed via “pip install multimatch”. A corresponding publication (Wagner & Hanke) in the Journal of Open Source Software (JOSS) is currently in preparation. The sourcecode can be found at github.com/AdinaWagner/multimatch.

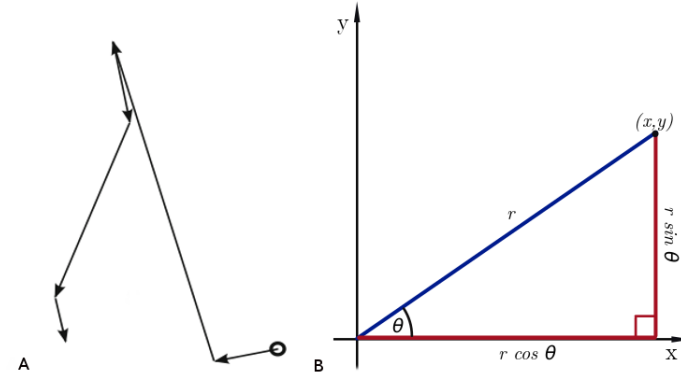


Figure 2.1: *A: Geometric scanpath representation.* An idealized saccade is represented as the shortest distance between two fixations. The Cartesian coordinates (x, y) of the fixations are thus the starting and ending points of a saccade. The length of a saccade in x (or y) direction is computed as the difference in x (or y) coordinates of starting and ending point.

B: Length and angle computation. Length from the coordinate origin, rho, is computed as the Euclidean norm by means of the Pythagorean theorem: $r = \sqrt{x^2 + y^2}$. The angle in radians, theta, is computed as a variation of the arctangent function as the angle from positive x axis to the point (x, y), with positive values denoting counterclockwise angle from positive x axis and negative values denoting clockwise angles: $\theta = \arctan2(x, y)$.

To reduce complexity, the scanpaths are simplified according to angle and amplitude in an iterative procedure. Two or more saccades are grouped together if angles between two consecutive saccades are below an angular threshold $TDir$, or if the amplitude of successive saccades is below a length threshold $TAmp$, as long as intermediate fixations of the saccades are shorter than a duration threshold, $TDur$. As such, small, locally contained saccades, and saccades in the same general direction are summed to form larger, less complex saccades.

In order to find pairings of saccade vectors to compare, the simplified scanpaths are temporally aligned. The aim is to not necessarily align two saccade vectors that constitute the same component in their respective vector sequence, but those two vectors that are the most similar while still preserving temporal order. In this way, a stray saccade in one of the two scanpaths does not lead to an overall low similarity rating, and it is further possible to compare scanpaths of unequal length. To do so, all possible pairings of saccades are evaluated in similarity by the vector differences between all pairings (shape): The vector differences between each element i in scanpath $S1 = u_1, u_2, \dots, u_m$ and each element j in scanpath $S2 = v_1, v_2, \dots, v_n$ are computed and stored in a matrix M with dimension $m \times n$ as weights that denote similarity, with low weights corresponding to high similarity. In a next step, an adjacency matrix is build, defining the rules on which connection between matrix elements are allowed to preserve the temporal order of saccades: In order to take temporal sequence of saccades into account, connections can only be made to the right, below or below-right (green arrows in figure 2.2). Together, matrices M and the adjacency matrix constitute a

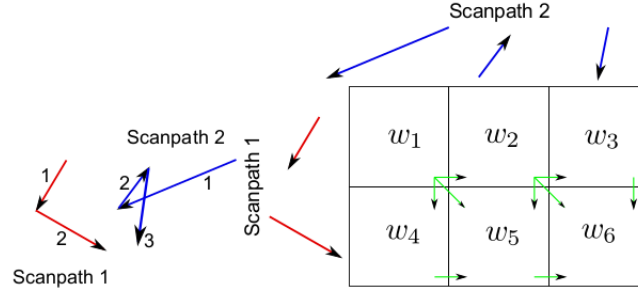


Figure 2.2: The elements of two hypothetical scanpaths (left) are used to compute vector differences between all possible pairings as weights w_1, \dots, w_6 in a matrix M (right). Green arrows indicate connection rules defined by an adjacency matrix. Taken from Jarodzka, Holmqvist, and Nyström (2010)

matrix representation of a directed, weighted graph (figure 2.2). Elements of the matrix are nodes, the connection rules constitute edges and the weights define the cost associated with each connection. In the generic example in figure 2.2 the edge between node (1, 1) and node (1, 2) has an associated weight, or cost, of w_2 .

A Dijkstra algorithm (Dijkstra 1959) is used to find the shortest path from the top left node, the first two saccade vectors, to the bottom right node, the last two saccade vectors. “Shortest” path is defined as the connection between nodes with the lowest possible sum of weights, i.e. the highest similarity. The path returned by the Dijkstra algorithm is a sequence of indexes, denoting pairings of saccade vectors from each scanpath, and as such the desired alignment of scanpaths (Dewhurst et al. 2012). Finally, in a last step, five measures of scanpath similarity (see figure 2.3) are computed for a multidimensional similarity evaluation. This is done by performing simple vector arithmetic on all aligned saccade pairs (u_i, v_j) , normalizing the results to range $[0, 1]$ according to a certain metric, and taking the median of the results. Higher values indicate higher similarity between scanpaths on the given dimension.

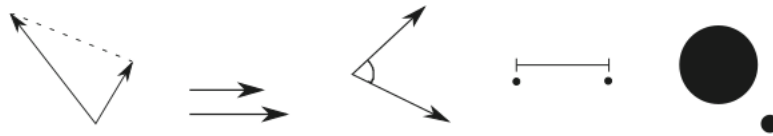


Figure 2.3: **Shape**: Vector difference between aligned scanpaths, normalised by 2x the screen diagonal. **Length**: Difference in vector lengths, normalized by the screen diagonal. **Direction**: Angular distance between saccade vectors, normalized by π . **Position**: Position difference of fixation, normalized by the screen diagonal. **Duration**: Duration differences of aligned fixations, normalized against maximum duration within the comparison.

Algorithm 1: multimatch

Input : n x 3 fixation vectors (x, y, duration), simplification thresholds T_{Amp} , T_{Dir} , T_{Dur}

Output: eventfile: onset of scanpath, duration, and five similarity measures range [0, 1]

```

1 for each fixvector do
2   for each fixation f in fixvector do
3     fixationx, fixationy, fixationdur = x, y, duration ;
4     saccadex(y) = fixationx(y)f - fixationx(y)f+1 ;
5     saccadelenx(leny) = fixationx(y) - saccadex(y);
6     saccaderho, saccadetheta = cartesian2polar(saccadelenx, saccadeleny) ;
7     eyedata = [fixationx, fixationy, fixationdur, saccadex, saccadey,
      saccadelenx, saccadeleny, saccaderho, saccadetheta]
8   end
9   while simplification is still possible do
10    for each saccade i in eyedata do
11      if angle(saccadei, saccadei+1) < TDir and fixationdur < TDur then
12        combine saccades
13      end
14      if saccaderho i, saccaderho i+1 < TAmp and fixationdur < TDur then
15        combine saccades
16      end
17    end
18  end
19 end
20 for each pair of (simplified) scanpaths  $S1 = \{u_1, \dots, u_m\}, S2 = \{v_1, \dots, v_n\}$  do
21   calculate M: Matrix of saccade length differences between each  $i \in S1$  and  $j \in S2$ ;
22   create a directed graph with elements of M as weights  $w_1, \dots, w_i$ ;
23   align scanpaths with  $\min(\sum w)$  from  $M(0, 0)$  to  $M(n, m)$  via Dijkstra algorithm;
24   add aligned scanpath to set of aligned scanpaths ;
25   for each pair( $u_i, v_i$ ) in aligned scanpaths do
26     angular difference = angle( $u_i, v_i$ );
27     position difference =  $\sqrt{(x(u_i) - x(v_i))^2 + (y(u_i) - y(v_i))^2}$  ;
28     vector difference =  $\sqrt{(\text{len}_x(u_i) - \text{len}_x(v_i))^2 + (\text{len}_y(u_i) - \text{len}_y(v_i))^2}$ ;
29     length difference = abs( $\text{rho}(u_i) - \text{rho}(v_i)$ );
30     duration difference = abs( $\text{dur}(u_i) - \text{dur}(v_i)$ ) ;
31   end
32   normalize angular difference by  $\pi$ ;
33   normalize vector difference by 2x the screen diagonal (maximum theoretical distance);
34   normalize length difference, position difference by the screen diagonal;
35   normalize duration difference by maximal duration difference in the scanpaths;
36   return median for each similarity measure return similarity measures
37 end

```

TODO: Not sure whether the pseudo code makes anything any clearer. Looks neat though and took a while. Might ban it into Appendix.

2.2.2 multimatch usage in the studyforrest dataset

The extraction of eye events for further use with multimatch to estimate a measure of exogenous attentional control included several considerations based on findings and known caveats in eye-tracking and attention research with dynamic scenes. The resulting additional functions are mainly concerned with automatic scanpath extraction from the approximately 15 minutes spanning and several eye-movement categories containing eye-tracking event files by user-defined rules. An overview in pseudocode can be found in algorithm 2.

First, as the stimulus was dynamic, the start and end of pursuit movements were relabeled as fixations. This was done to accommodate the fact that most areas of interest in a movie are in motion. Focusing on the running young Forrest for example would likely appear as a slow pursuit in the eye event data. Disregarding such pursuits would lead to a large loss in data and variability, and as such, they were included via relabeling.

A different consideration concerned the selection of scanpaths from the approximately 15 minute long segments, as they are unsuitably long to derive a continuous attention measure or do a scanpath comparison on. It has been shown that subjects gazes have a bias towards the center in Hollywood movies (Tseng et al. 2009). This bias can at least in part be traced back to a strong center bias directly after cuts in dynamic scenes. Carmi and Itti (2006) found the highest degree of exogenous attentional control (i.e. maximal inter-observer similarity in gaze as recorded via eye-tracking) in a moving collage of 50 heterogeneous video clips directly after cuts to new, semantically unrelated scenes. Without isolating cuts it therefore can not be disentangled whether the contribution of visual features of the movie on scanpaths stems from exogenous control from a movie element or from the sudden occurrence of a new scene after a cut. To not introduce such a confound in the eye-tracking data, the data was cut into segments that did not contain cuts to different scenes, relying on a location annotation by Häusler and Hanke (2016). They define a *shot* as a movie sequence between two cuts. The location annotation contained a detailed description of the timing of each of the 870 shots of the movie as well as information about the depicted location in different levels of abstraction. In a first step, all consecutive short shots to the same locale (e.g. two consecutive, short movie shots in Forrest’s bedroom, without any cuts containing semantically unrelated content) were grouped together. A “short” shot was defined as being shorter than the median length of 4.92s. This procedure increased the median length of shots from 4.92s to 7.019s. These elongated shots are henceforth referred to as *snippets*. In a next step, within the scanpath comparison, the onset and offset times of the resulting snippets were extracted for every snippet longer than 4.92s. To not introduce an influence of snippet length into the similarity results, as longer movie snippets will be more likely to be less similar than short ones, scanpath lengths were standardized to approximately 4.92s seconds (the original median length). To further evade any problems associated with the center bias, scanpaths were extracted from the end of the snippet: The last oculomotor event within the range of the snippet marked

the end of a scanpath. As such, scanpaths began maximally distant to the snippet onset. As the eye movements of participants are interindividually different and do not correspond exactly to snippet timing, the precise onsets and resulting exact durations were computed as well, and stored for later use as regressors.

Algorithm 2: The studyforrest specific functions of multimatch

Input : REMoDNaV eye movement datafiles per run, movie location annotation per run, desired length to elongate shots to $ldur$, desired scanpath length dur

Output: One $n \times 3$ fixation vector (x, y, duration) for any snippet longer than dur

```

1 if  $event_{row} = pursuit$  then
2   | regard start and end points of pursuit as fixation
3 end
4 for each row in REMoDNaV datafile do
5   | if  $event_{row} = fixation$  then
6   |   | extract x-, y-coordinate, duration
7   | end
8   | if  $x, y < 0$  or  $x > 1280$  or  $y < 720$  then
9   |   | discard as out-of-bound gaze
10  | end
11 end
12 for each row in annotation do
13   | if  $(locale_{row} = locale_{row+1})$  and  $(duration_{row}, duration_{row+1} < ldur)$  then
14   |   | combine to one shot
15   | end
16 end
17 for each row in annotation do
18   | if  $duration_{row} > dur$  then
19   |   | extract shotonset and offset time
20   | end
21 end
22 for each  $i$  in  $length(onset)$  do
23   |  $fixvector_i = REMoDNaV[REMoDNaV_{onset} > onset_i; REMoDNaV_{onset} < offset_i]$ 
24 end
25 return  $\{fixvector_1, \dots, fixvector_i\}, onsets, durations$ 

```

TODO: Not sure whether the pseudo code makes anything any clearer. Looks neat though and took a while. Might ban it into Appendix.

2.2.3 Deriving a measure of attentional modulation from multimatch

A common measure of attentional modulation was computed in a two-step procedure. First, scanpath comparisons of all scanpaths from the same shot of two subjects were calculated for all possible pairs of subject. This resulted in $C_k(n) = \binom{N}{k} = \binom{15}{2} = 105$ combinations for $N = 15$ subjects. These comparisons were done without any further simplification (i.e. no use of the direction, length, and duration thresholds), as

even minor differences in scanpaths obtained from a movie can correspond to major differences in attended visual stimuli. In a second step, the resulting similarities for each of the five similarity dimensions were averaged. Thus, for each snippet longer than 4.92s five similarity measures were computed that represented the average similarity of scanpaths of all subjects on the given dimension.

2.3 A novel approach for the analysis of ROI specificity

This section outlines the proposed method for ROI specificity analysis in general, and provides exemplary use cases in the form of two sequential preliminary validation analysis, independent of the FEF.

The general idea behind the proposed method is to derive a functional signature as a temporal course from two ROIs in question, and to explain the functional signature by providing a detailed description of the experimental design. This is realized with a two step procedure consisting of (1) classification and (2) GLM analysis. Simultaneously, the method relies on transposing the fMRI dataset, such that all time points are features, and the voxels are samples.

The functional signature is obtained from a classifiers' sensitivity profile in an n-fold leave-one-participant-out cross-validation: The classifier is trained on all subjects except for one and then tested on that out-of-sample individual, and this method is repeated for each of the subjects (step 1, classification analysis). During classification, feature weights are obtained for the decision between all pairwise combinations of classes. Obtained sensitivity profiles per fold are averaged across folds. As a result, there is one temporal sensitivity profile for each pairwise ROI combination. The achieved overall classification accuracy serves as a measure of whether population wide functional specificity and sensitivity between the ROIs can be captured. In the second step (GLM), the averaged sensitivity profile is regressed onto the stimulus description for an estimate on how well the experimental design could describe ROI differences, and to point to the effects in the experimental design contributing to an ROI distinction as evident from the beta coefficients of the respective regressors.

The classifier of choice, a Gaussian Naive Bayes classifier, was readily available in pyMVPA's (Hanke et al. 2009) `mvpa2.suite` python module as `GNB()` and customizable to be a linear classifier with the parameter `common_variance = True`. However, it lacked functionality to compute feature weights for a sensitivity analysis. Therefore, as a preliminary step, the `GNB` class in the sourcecode of the `pymvpa` package was extended with the implementation of the subclass `GNBWeights`. This made a sensitivity analysis possible, that returns the weights $w_i = \sum_i \frac{\mu_{i,0} - \mu_{i,1}}{\sigma_i^2}$ for each feature for any possible pair of labels in a binary decision. For an example, consider a classification task with three labels L0, L1, L2, and two features, F1, F2. The implemented sensitivity analyzer will for each possible pair of labels (L0 versus L1, L0 versus L2, L1 versus

L2) return one weight per feature⁶.

As the datasets need to be transposed to have time points as features, each weight corresponded to the difference in mean activation of all voxels given the class at one time point, normalized by their common variance. In case of for example the decision between FFA and PPA, a negative weight therefore corresponds to an overall higher PPA activation at this particular point in time. The type of stimulation present at this point in the experimental design likely contributes to the distinction between the two areas, and the magnitude of its influence will be visible in its associated beta coefficient obtained from the GLM.

2.3.1 Method validation

In order to establish feasibility of the procedure and to test the generalization of the developed approach, the method was validated in a two-step procedure. As a test of plausibility, it was used on data from a common block design for ROI localization within the studyforrest dataset. As a measure of its capabilities to generalize to naturalistic stimulation, it was subsequently used with the studyforrests' movie data, collected on the same participants. For both datasets, six functional ROI masks (FFA, PPA, OFA, EBA, LOC, early visual cortex) were available to serve as labels during classification analysis. Both validating analyses were restricted to the FFA-PPA distinction as for the wealth of literature concerned with these ROIs.

For use in the localizer dataset, the event files describing the stimulation protocol of the object-category task were used to compute a group-level event file. This was possible because the order of blocks was the same across runs and participants (see Sengupta et al. 2016). For each participant, the onsets of each image were averaged. Deviations from individual onsets to the average group onsets were smaller than 0.5 seconds. In addition to the six object categories, the first occurrence of an image within an object category was classed a new type of event with the suffix 'first', resulting in 12 regressors (face, face_first, body, body_first, house, house_first, scene, scene_first, object, object_first, scramble, scramble_first). This was done to account for any additional variation introduced by the change in categories. Using pyMVPA's `fit_event_hrf_model()` function, the resulting events were entered as regressors to explain the time course of sensitivities.

For generalization in the movie dataset, a stimulus description of the movie was obtained from two sources, the location annotation of Häusler and Hanke (2016), and an automatic facial feature extraction per frame of Hanke (unpublished, computed using Google's Cloud Vision API via pliers (McNamara et al. 2017)). An event was derived for each "setting" in the location annotation occurring more than once. An event was further derived for night-time, a scene being located exterior, a general cut, and contextual jumps to future or past from the previous shot. The facial feature data

⁶The sourcecode for this change can currently be found in the following PR: <https://github.com/PyMVPA/PyMVPA/pull/591/files>

was downsampled to 1 second and two conditions were derived, "face" for any 1 second time window with a maximum of less than 3 faces, and "many_faces" for any 1 second time window with more faces present.

2.4 TODO: FEF

[this is a gigantic TODO]

add something in the notion of: Study of the frontal eye fields attempting to capture its involvement as both a structure of motion and attentional deployment ... somewhere

3 Results

3.1 Eye movements and multimatch

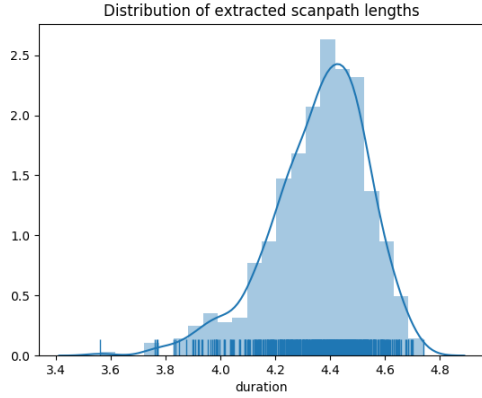
In order to check the validity of the eye-movement extraction with REMoDNaV, the mean and median duration of extracted fixations were compared to the results of Dorr et al. (2010). Fixations as found by REMoDNaV had a length of mean = 0.326s, median = 0.256s, which corresponds closely to the aforementioned study’s results (0.354s and 0.253s, respectively). A comparison of algorithm performance and human coding of eye movements on data of Andersson et al. (2017) by **Dar et al. (in preparation)** showed the highest degree of concordance of all contemporary algorithms used in Anderson et al. (2015), providing further evidence for the validity of extracted eye movements.

Based on the REMoDNaV output, the multimatch method was able to extract a total of $N = 533$ scanpaths from the movie. The median duration of extracted scanpath duration was 4.39 seconds (mean = 4.36s) (see figure 3.1). The median and average similarities per dimension are stated in table 3.1.

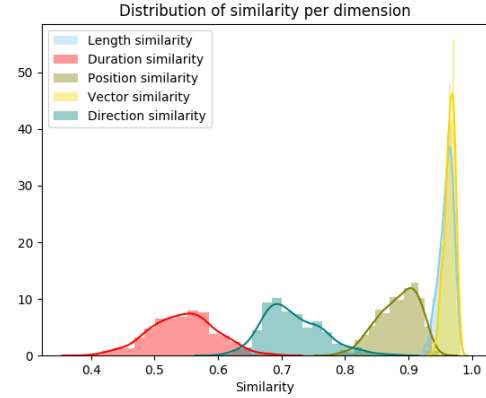
Variable	mean [SD]	median
Vector similarity	0.97 [0.01]	0.97
Position similarity	0.88 [0.03]	0.89
Length similarity	0.96 [0.01]	0.96
Duration similarity	0.54 [0.05]	0.55
Direction similarity	0.72 [0.05]	0.71

Table 3.1: Characteristics of scanpath similarity across the movie.

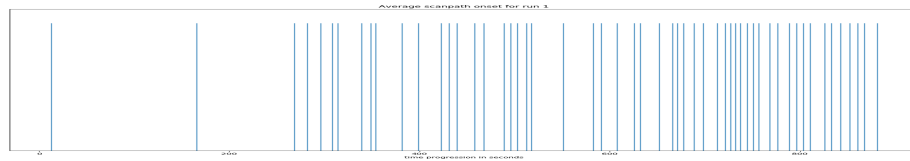
Figure 3.1 gives an overview of some additional scanpath characteristics, in particular the distribution of similarity per dimension, the onset of scanpaths in the first run, and an example of movie segments with high and low similarity.



(a) Extracted scanpath lengths



(b) Distribution of similarities per dimension for the movie



(c) Average onsets of scanpaths in seconds for the first run



(d) Frame from movie snippet with the lowest average similarity rating of run 1, with overlaid eye-gaze



(e) Frame from movie snippet with the highest average similarity rating of run 1, with overlaid eye-gaze

Figure 3.1: Overview of extracted scanpaths' characteristics.

TODO:learn inkscape an arrange it nicely

3.2 Method validation

Figure 3.2 shows the time course of sensitivities and the corresponding GLM fit for the localizer experiment (top panel A) and the movie data (top panel B) for the first run of each of these experiments. For the localizer data, 89% of the variance in the sensitivity time course was explained with the stimulation protocol description. For the movie data, 28% of the variance was explained with a general location and face-occurrence annotation. **TODO: for now, don't include classification accuracies, they are irrelevant here**

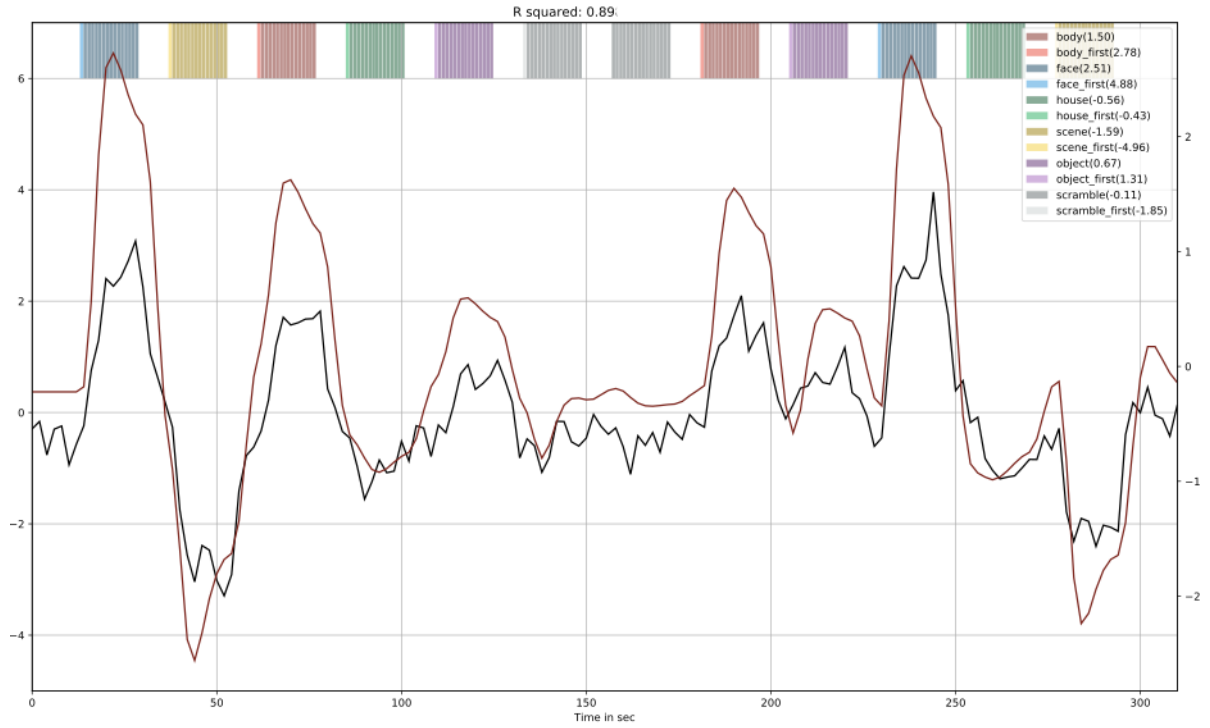
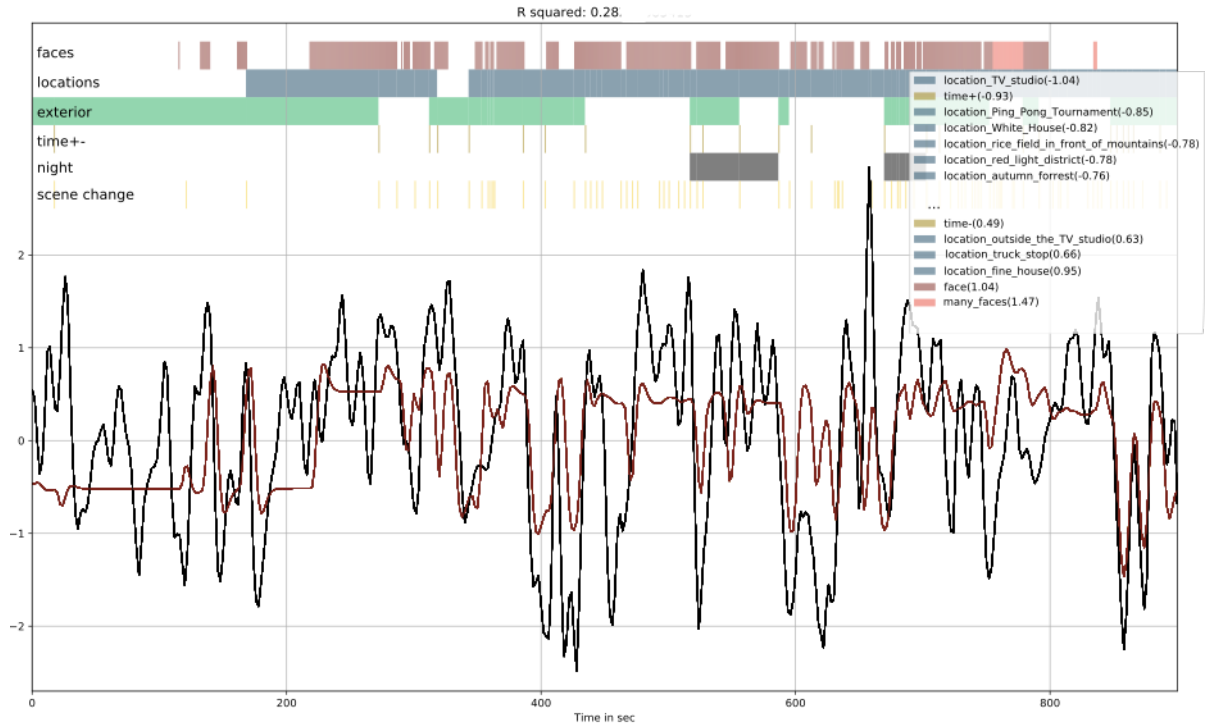
A. Localizer**B. Movie**

Figure 3.2: Black line: Sensitivities. Red line: GLM fit from regressing sensitivities onto the event regressors. A: Run 1 of the localizer experiment, with 12 regressors derived from the object categories of the stimulation. B: Run 1 of the movie experiment with regressors derived from the movie annotation.

Colors correspond to stimulation with one particular event type in time, beta values are given in brackets. The legend in panel B contains only the diagnostically most valuable regressors

3.3 TODO FEF

large TODO

4 TODO Discussion

do a short summary of everything

- The aim of this thesis was to...
- Therefore, in this thesis ... What did I do: MultiMatch, functional specificity analysis
- The results show that...

MultiMatch

The implemented multimatch algorithm was able to extract scanpaths and calculate their similarity on five different dimensions. The average length of extracted scanpaths was shorter than the duration aimed at (4.92s), as the onset and offset of the required eye movements did rarely correspond exactly to the onsets and offsets of the selected movie snippet. Nevertheless, the extracted scanpaths were similar in duration, and hence do not introduce a bias of scanpath lengths into the similarity calculation. With the exception of the movie start with three very long shots, the extracted scanpaths are distributed evenly across the movie. As evident from table 3.1 and figure 3.1, scanpaths were almost perfectly similar on the dimensions vector length and vector position. This is likely at least partially due to the scanpath alignment based on the scanpath shape. Scanpaths were also highly similar on the position dimension, which underlines the strong gaze control of the movie stimulus. Subjects scanpaths differed more substantially on the dimensions direction and duration, which indicates differences in fixation dwelling times and saccadic angle. Thus, the general points of interest (as evident from high similarities in position, length and shape) were similar across subject, but differences in direction and duration might indicate interindividually different exploration strategies. All dimensions show a remarkable consistency in similarity measures as evident from the small standard deviations. Following the reasoning behind the implementation of the method, this might indicate a consistently high level of exogenous control by the movie stimulus. This finding is consistent with research on viewing behavior during movies: Unlike during static image viewing, the spatio-temporal gaze behavior of multiple viewers exhibits a substantial degree of coordination in movie watching. Smith and Henderson (2008) coined the term *attentional synchrony* for this phenomenon. During attentional synchrony, viewers gazes cluster around a small portion of the screen at any one moment. Goldstein, Woods, and Peli (2007), for example, found the distribution of fixations of viewers to occupy less than 12% of the total screen area in more than 50% of the time in six Hollywood movies. In

a comparison between different types of static and dynamic visual stimuli, Dorr et al. (2010) found the highest consistency between viewers' eyegazes during professionally produced (Hollywood) movies, likely largely due to the use of cinematic composition of scenes, deliberate camera work and editing. Hasson, Yang et al., 2008 found high correspondence in gaze behavior across subjects, even for backwards presentations of movies.

Functional specificity analysis

The results of the validation analysis of the functional specificity analysis provide a general proof of principle for the method. It achieved a reasonable fit to the sensitivity time course, both for the localizer and the movie data (figure 3.2), given the comprehensiveness of the used regressors: In the localizer paradigm, the stimulation protocol is able to explain 89% of the variance in the sensitivity profile. As visible both from the time series' appearance and the regression coefficients, stimulation with images of faces ($\beta = 2.51$) and, to a lesser extent, bodies ($\beta = 1.50$), in particular the first of these stimuli per block ($\beta=4.88$ and $\beta=2.78$, respectively), differentially activated FFA more than PPA. PPA in turn was differentially activated by stimulation with scenes (scene_first: $\beta = -4.96$, scene = $\beta = -1.59$). These results are in accordance to the literature (e.g. Fox, Iaria, and Barton 2009, ...more here...) → proof of principle.

For the movie data, the method achieved a fit of $R^2 = .28$. The functional description of the movie was hence only able to explain less than a third of the variance in the activation differences between FFA and PPA over the course of the movie. Nevertheless, this fit was achieved with a rather coarse descriptions of a shots setting without any additional filtering, and a general face annotation, over a time span of roughly 2 hours. Moreover, the resulting regression coefficients are largely in line with previous results. The FFA is stronger activated than the PPA in the many_faces ($\beta=1.47$) and face ($\beta=1.04$) conditions. **TODO: lets cut this here, and put the rest into future directions**

4.1 TODO Drawbacks and shortcomings

- A number of potential issues need to be addressed to access the validity of the results (maybe refer to Deborah Mayors Severe Testing ideas?)
- As evident from the consistency in scanpath similarity, the MultiMatch method was not able to detect large changes in attentional mode throughout the movie. This might reflect a consistently high exogenous control exercised by the movie stimulus. However, it is not clear whether similarity in scanpaths does actually correspond to exogenous control. The similarity measures obtained from eye gazes could fail to account to differentiate exogenous modulation from endogenous influences arising from a shared goal between participants. Consider the following example: in order to fully comprehend the movie content, it might be of relevance for every subject to endogenously focus on the facial features of Forrest in his

dialogue with Jenny. This scene does not contain salient stimulus features or other types of exogenous control, yet it would lead to a high similarity measure. Another possible confound is an endogenous attentional control evoked by the narration in the overlaid audio-description. The narrative voice describing the visual contents of the scenes could shift the viewers attention in an endogenous, yet inter-individually highly correlated fashion. A high degree of synchrony in eye movements therefore would not necessarily correspond to purely exogenous influences. A more thorough investigation of the origin of eye gaze similarity might lie in a subsequent qualitative analysis of individual movie shots, but that was not within the scope of this thesis.

- Problem of 5 different similarity measures — BUT successfully used by French et al., multidimensional construct preferable as might explain more variance (citation? check anderson or the other dewhurst publication)
- The thesis was largely exploratory in regard to determining the functional specificity of ROIs. Of course, the value and novelty of the developed method lies in its ability to explore previously unknown functions with naturalistic stimulation. As such, it is providing interesting new insights on the level of functional neuroanatomy, and constitutes additional evidence that the use of naturalistic stimulation is of immense value in neuroimaging research and can underlie experimental control (Hasson et al. 2004). Nevertheless, any exploratory analysis is lacking a thorough derivation of hypothesis from theory, and the preregistration of such hypotheses to guarantee scientific integrity. Furthermore, given that the choices of annotation for naturalistic stimulation and their inclusion can potentially be unlimited, the method leaves many researchers degrees of freedom to potentially bend results into shape. Therefore, the method does not suffice on its own. Instead, it should be regarded as a valuable first step to gain exploratory insights and generate novel hypotheses. The ideas derived from the method should subsequently be brought to test on new data to prevent findings due to chance being published without further testing.
- Address the question of whether the linearity and gaussian distribution assumptions hold.
- Now trash all shortcomings and come to the conclusion that for was done and despite a number of limitations this thesis was able to advance the literature in the following ways

4.2 TODO Future directions

- The employed method for determining the functional specificity of ROIs proved to be an interesting novel approach to gain further insights into the functional neuroanatomy of regions of interest. However, the achieved model fit was not

optimal. A follow-up project of this thesis will explore the benefit of a different classification algorithm, a *support vector machine* (SVM), and continue to explore the insights of this method in regard to a studyforrest movie dataset. Furthermore, an additional improvement to make the classification more robust could be to incorporate the voxels proximity to blood vessels (in light of the "draining vein effect"¹). The studyforrest dataset might contain appropriate angiographic data from phase 1 to pursue this enhancement.

- it will be especially interesting to see whether the method is capable of providing insights into functional specificities of the same structure in different hemispheres, for example the left and the right FFA.
- outline more promising research questions

4.3 TODO Conclusions

- now what have we all learned after reading this far?
- surely something needs additional research

Independent of the results, this thesis provided an open source implementation of the matlab based MultiMatch toolbox for multidimensional scanpath comparison. This might enable additional research into scanpath similarity with this geometric, multidimensional approach, without requiring usage and purchase of closed source software. Furthermore, it provided a proof of principle that the sensitivities of a "transposed" approach of classification in unison with a functional description of the experimental design can be used to gain insights into the functional neuroanatomy of regions of interest in a meaningful way. The work undertaken in this thesis constitutes a starting point for further development and future application of the method.

¹see <http://www.thebrainblog.org/2018/12/23/twenty-six-controversies-and-challenges-in-fmri/>

5 Things that are still shitty

big shit

- introduction is somewhat weak and lacks a common thread. thats because i lack a common thread though. will be easier to work on once the rest is finished.
- surely incomprehensible method description, especially ROI specificity analysis
- it's worthwhile to think about splitting the thesis from intro-methods-results-discussion into intro-methods1-results1-methods2-results2-generaldiscussion

small shit

- too many passive constructions
- many long, incomprehensible sentences. boil it down to something simpler
- I need better verbal connections between sections. Short intros and outros of all main chapters.
- Citations are fucked up, take care of the bibtex style, its still according to macroeconomics, there has to be an example for APA on the internet somewhere
- all **bold** citations are not yet in bibtex, mostly because they are weird formats.
- prune references extensively after recent shortening spree (maybe simply disable nocite* if I'm lazy?)
- literally just barfed every single abbreviation into the list of abbreviations, I doubt this is ultimately useful. Also still contains GLM abbreviations from vwl
- Things are not optimally aranged currently, once the text is in final form make sure to force figures and algorithms to their appropriate places.
- think of better short captions for things in the lists of figures and tables.

Bibliography

- Alkan, Y., B. B. Biswal, and T. L. Alvarez (2011). “Differentiation between vergence and saccadic functional activity within the human frontal eye fields and midbrain revealed through fMRI.” *PLoS One* 6.11, e25866.
- Amiez, C. and M. Petrides (2009). “Anatomical organization of the eye fields in the human and non-human primate frontal cortex.” *Progress in neurobiology* 89.2, pp. 220–230.
- Anderson, N. C. et al. (2015). “A comparison of scanpath comparison methods.” *Behavior research methods* 47.4, pp. 1377–1392.
- Andersson, R. et al. (2017). “One algorithm to rule them all? An evaluation and discussion of ten eye movement event-detection algorithms.” *Behavior research methods* 49.2, pp. 616–637.
- Bellebaum, C., P. Thoma, and I. Daum (2012). *Neuropsychologie*. Springer.
- Blanke, O. et al. (2000). “Location of the human frontal eye field as defined by electrical cortical stimulation: anatomical, functional and electrophysiological characteristics.” *NeuroReport* 11.9, pp. 1907–1913.
- Blanke, O. and M. Seeck (2003). “Direction of saccadic and smooth eye movements induced by electrical stimulation of the human frontal eye field: effect of orbital position.” *Experimental brain research* 150.2, pp. 174–183.
- Blankenburg, F. et al. (2010). “Studying the role of human parietal cortex in visuospatial attention with concurrent TMS–fMRI.” *Cerebral Cortex* 20.11, pp. 2702–2711.
- Bressler, S. L. et al. (2008). “Top-down control of human visual cortex by frontal and parietal cortex in anticipatory visual spatial attention.” *Journal of Neuroscience* 28.40, pp. 10056–10061.
- Bruce, C. J. et al. (1985). “Primate frontal eye fields. II. Physiological and anatomical correlates of electrically evoked eye movements.” *Journal of neurophysiology* 54.3, pp. 714–734.
- Carmi, R. and L. Itti (2006). “Visual causes versus correlates of attentional selection in dynamic scenes.” *Vision research* 46.26, pp. 4333–4345.
- Carrasco, M. (2011). “Visual attention: The past 25 years.” *Vision research* 51.13, pp. 1484–1525.
- Chica, A. B., P. Bartolomeo, and J. Lupiáñez (2013). “Two cognitive and neural systems for endogenous and exogenous spatial attention.” *Behavioural brain research* 237, pp. 107–123.
- Connolly, J. D. et al. (2002). “Human fMRI evidence for the neural correlates of preparatory set.” *Nature neuroscience* 5.12, p. 1345.

- Corbetta, M., G. Patel, and G. L. Shulman (2008). “The reorienting system of the human brain: from environment to theory of mind.” *Neuron* 58.3, pp. 306–324.
- Corbetta, M. and G. L. Shulman (2002). “Control of goal-directed and stimulus-driven attention in the brain.” *Nature reviews neuroscience* 3.3, p. 201.
- Coutrot, A., J. H. Hsiao, and A. B. Chan (2018). “Scanpath modeling and classification with hidden Markov models.” *Behavior research methods* 50.1, pp. 362–379.
- Cristino, F. et al. (2010). “ScanMatch: A novel method for comparing fixation sequences.” *Behavior research methods* 42.3, pp. 692–700.
- Crowne, D. P., C. H. Yeo, and I. S. Russell (1981). “The effects of unilateral frontal eye field lesions in the monkey: visual-motor guidance and avoidance behaviour.” *Behavioural brain research* 2.2, pp. 165–187.
- Daniel, P. and D. Whitteridge (1961). “The representation of the visual field on the cerebral cortex in monkeys.” *The Journal of physiology* 159.2, pp. 203–221.
- Dewhurst, R. et al. (2012). “It depends on how you look at it: Scanpath comparison in multiple dimensions with MultiMatch, a vector-based approach.” *Behavior research methods* 44.4, pp. 1079–1100.
- Dijkstra, E. W. (1959). “A note on two problems in connexion with graphs.” *Numerische mathematik* 1.1, pp. 269–271.
- Dorr, M. et al. (2010). “Variability of eye movements when viewing dynamic natural scenes.” *Journal of vision* 10.10, pp. 28–28.
- Downar, J. et al. (2000). “A multimodal cortical network for the detection of changes in the sensory environment.” *Nature neuroscience* 3.3, p. 277.
- (2001). “The effect of task relevance on the cortical response to changes in visual and auditory stimuli: an event-related fMRI study.” *Neuroimage* 14.6, pp. 1256–1267.
- Drenckhahn, D. (2004). *Anatomie, Makroskopische Anatomie, Embryologie und Histologie des Menschen.: Band 2: Herz-Kreislauf-System, Lymphatisches System, Endokrine Drüsen, Nervensystem, Sinnesorgane, Haut.* Amsterdam: Elsevier GmbH.
- Fox, C. J., G. Iaria, and J. J. Barton (2009). “Defining the face processing network: optimization of the functional localizer in fMRI.” *Human brain mapping* 30.5, pp. 1637–1651.
- Fox, M. D. et al. (2006). “Spontaneous neuronal activity distinguishes human dorsal and ventral attention systems.” *Proceedings of the National Academy of Sciences* 103.26, pp. 10046–10051.
- French, R. M., Y. Gladys, and J.-P. Thibaut (2017). “An evaluation of scanpath-comparison and machine-learning classification algorithms used to study the dynamics of analogy making.” *Behavior research methods* 49.4, pp. 1291–1302.
- Goldstein, R. B., R. L. Woods, and E. Peli (2007). “Where people look when watching movies: Do all viewers look at the same place?” *Computers in biology and medicine* 37.7, pp. 957–964.
- Goodwin, J. (2007). “Cranial Nerves III, IV, and VI: the oculomotor system.” *Textbook of clinical neurology*, pp. 133–164.

- Gorgolewski, K. J. et al. (2016). “The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments.” *Scientific Data* 3, p. 160044.
- Gorgolewski, K. J. et al. (2017). *nipy/nipype: Release 0.13.1*. DOI: 10.5281/zenodo.581704. URL: <https://doi.org/10.5281/zenodo.581704>.
- Hanke, M. et al. (2009). “PyMVPA: a python toolbox for multivariate pattern analysis of fMRI data.” *Neuroinformatics* 7.1, pp. 37–53.
- Hanke, M. et al. (2014). “A high-resolution 7-Tesla fMRI dataset from complex natural stimulation with an audio movie.” *Scientific data* 1, p. 140003.
- Hanke, M. et al. (2016). “A studyforrest extension, simultaneous fMRI and eye gaze recordings during prolonged natural stimulation.” *Scientific data* 3, p. 160092.
- Hasson, U., R. Malach, and D. J. Heeger (2010). “Reliability of cortical activity during natural stimulation.” *Trends in cognitive sciences* 14.1, pp. 40–48.
- Hasson, U. et al. (2004). “Intersubject synchronization of cortical activity during natural vision.” *science* 303.5664, pp. 1634–1640.
- Hasson, U. et al. (2008). “Neurocinematics: The neuroscience of film.” *Projections* 2.1, pp. 1–26.
- Häusler, C. O. and M. Hanke (2016). “An annotation of cuts, depicted locations, and temporal progression in the motion picture "Forrest Gump".” *F1000Research* 5.
- He, B. J. et al. (2007). “Breakdown of functional connectivity in frontoparietal networks underlies behavioral deficits in spatial neglect.” *Neuron* 53.6, pp. 905–918.
- Holmqvist, K. et al. (2011). *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford.
- Ioannides, A. A. et al. (2004). “MEG tomography of human cortex and brainstem activity in waking and REM sleep saccades.” *Cerebral Cortex* 14.1, pp. 56–72.
- Itti, L. and C. Koch (2001). “Computational modelling of visual attention.” *Nature reviews neuroscience* 2.3, p. 194.
- Jarodzka, H., K. Holmqvist, and M. Nyström (2010). “A vector-based, multidimensional scanpath similarity measure.” *Proceedings of the 2010 symposium on eye-tracking research & applications*. ACM, pp. 211–218.
- Jenkinson, M. et al. (2012). “Fsl.” *Neuroimage* 62.2, pp. 782–790.
- Kawashima, R. et al. (1998). “Oculomotor sequence learning: a positron emission tomography study.” *Experimental brain research* 122.1, pp. 1–8.
- Kowler, E. (2011). “Eye movements: The past 25 years.” *Vision research* 51.13, pp. 1457–1483.
- Krauzlis, R. J. (2014). “Eye movements.” *Fundamental Neuroscience (Fourth Edition)*. Elsevier, pp. 697–714.
- Leichnetz, G. and M. Goldberg (1988). “Higher centers concerned with eye movement and visual attention: cerebral cortex and thalamus.” *Reviews of oculomotor research* 2, pp. 365–429.
- Levenshtein, V. I. (1966). “Binary codes capable of correcting deletions, insertions, and reversals.” *Soviet physics doklady*. Vol. 10. 8, pp. 707–710.

- Malinen, S., Y. Hlushchuk, and R. Hari (2007). “Towards natural stimulation in fMRI—issues of data analysis.” *Neuroimage* 35.1, pp. 131–139.
- Mitchell, T. M. (1997). *Machine learning*. McGraw Hill series in computer science. McGraw-Hill. ISBN: 978-0-07-042807-2. URL: <http://www.worldcat.org/oclc/61321007>.
- Monosov, I. E. and K. G. Thompson (2009). “Frontal eye field activity enhances object identification during covert visual search.” *Journal of neurophysiology* 102.6, pp. 3656–3672.
- Muggleton, N. G. et al. (2003). “Human frontal eye fields and visual search.” *Journal of neurophysiology* 89.6, pp. 3340–3343.
- Munoz, D. P. (2002). “Commentary: saccadic eye movements: overview of neural circuitry.” *Progress in brain research*. Vol. 140. Elsevier, pp. 89–96.
- Nyström, M. and K. Holmqvist (2010). “An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data.” *Behavior research methods* 42.1, pp. 188–204.
- Ouerhani, N. et al. (2004). “Empirical validation of the saliency-based model of visual attention.” *ELCVIA: electronic letters on computer vision and image analysis* 3.1, pp. 13–24.
- Paus, T. (1996). “Location and function of the human frontal eye-field: a selective review.” *Neuropsychologia* 34.6, pp. 475–483.
- Perry, V and A Cowey (1986). “The ganglion-cell and cone distributions in the monkeys retina - Implications for central magnification factors.” *Behavioural Brain Research* 20.1, pp. 131–132.
- Petit, L. and J. V. Haxby (1999). “Functional anatomy of pursuit eye movements in humans as revealed by fMRI.” *Journal of Neurophysiology* 82.1, pp. 463–471.
- Poldrack, R. A., Y. O. Halchenko, and S. J. Hanson (2009). “Decoding the large-scale structure of brain function by classifying mental states across individuals.” *Psychological science* 20.11, pp. 1364–1372.
- Posner, M. I., C. R. Snyder, and B. J. Davidson (1980). “Attention and the detection of signals.” *Journal of experimental psychology: General* 109.2, p. 160.
- Ross, N. M. and E. Kowler (2013). “Eye movements while viewing narrated, captioned, and silent videos.” *Journal of vision* 13.4, pp. 1–1.
- Ruff, C. C. et al. (2009). “Hemispheric differences in frontal and parietal influences on human occipital cortex: direct confirmation with concurrent TMS–fMRI.” *Journal of Cognitive Neuroscience* 21.6, pp. 1146–1161.
- Schall, J. D. et al. (1995). “Topography of visual cortex connections with frontal eye field in macaque: convergence and segregation of processing streams.” *Journal of Neuroscience* 15.6, pp. 4464–4487.
- Sengupta, A. et al. (2016). “A studyforrest extension, retinotopic mapping and localization of higher visual areas.” *Scientific data* 3, p. 160093.
- Shulman, G. L. et al. (2003). “Quantitative analysis of attention and detection signals during visual search.” *Journal of neurophysiology* 90.5, pp. 3384–3397.

- Shulman, G. L. et al. (2007). “Right TPJ deactivation during visual search: functional significance and support for a filter hypothesis.” *Cerebral Cortex* 17.11, pp. 2625–2633.
- Siegelbaum, S. A. and A. J. Hudspeth (2000). *Principles of Neural Science (Kandel)*. McGraw Hill Education Medical. New York: McGraw-Hill. ISBN: 978-0071390118.
- Smith, T. and J. Henderson (2008). “Attentional synchrony in static and dynamic scenes.” *Journal of Vision* 8.6, pp. 773–773.
- Smith, T. J. (2013). “Watching you watch movies: Using eye tracking to inform film theory.”
- Stanton, G., C. Bruce, and M. Goldberg (1995). “Topography of projections to posterior cortical areas from the macaque frontal eye fields.” *Journal of Comparative Neurology* 353.2, pp. 291–305.
- Tehovnik, E. J. et al. (2000). “Eye fields in the frontal lobes of primates.” *Brain Research Reviews* 32.2-3, pp. 413–448.
- Thompson, K. G. and N. P. Bichot (2005). “A visual salience map in the primate frontal eye field.” *Progress in brain research* 147, pp. 249–262.
- Tseng, P.-H. et al. (2009). “Quantifying center bias of observers in free viewing of dynamic natural scenes.” *Journal of vision* 9.7, pp. 4–4.
- Vernet, M. et al. (2014). “Corrigendum: Frontal eye field, where art thou? Anatomy, function, and non-invasive manipulation of frontal regions involved in eye movements and associated cognitive operations.” *Frontiers in integrative neuroscience* 8, p. 88.
- Vogt, N. (2018). “Machine learning in neuroscience.” *Nature Methods* 15.1, p. 33.
- Vossel, S., J. J. Geng, and G. R. Fink (2014). “Dorsal and ventral attention systems: distinct neural circuits but collaborative roles.” *The Neuroscientist* 20.2, pp. 150–159.
- Weintraub, S. and M.-M. Mesulam (1987). “Right cerebral dominance in spatial attention: Further evidence based on ipsilateral neglect.” *Archives of neurology* 44.6, pp. 621–625.
- Yarbus, A. L. (1967). “Eye movements during perception of complex objects.” *Eye movements and vision*. Springer, pp. 171–211.
- Zeigermann, O. and C. Nguyen (2018). *Machine Learning – kurz & gut: Eine Einführung mit Python, Pandas und Scikit-Learn*. kurz & gut. O’Reilly. ISBN: 9783960101932. URL: <https://books.google.com/books?id=n-FVDwAAQBAJ>.

Appendices

What needs to go in here?

- maybe code? though contains links to github already
- correlation tables?