Kaisu Ölander

Eye movements and early magnetoencephalographic brain responses induced by faces in natural scenes

School of Electrical Engineering

Thesis submitted for examination for the degree of Master of Science in Technology.

Espoo 30.10.2015

Thesis supervisor:

Prof. Riitta Hari

Thesis advisor:

D.Sc. Linda Henriksson

Aalto University
School of Electrical Engineering

AALTO UNIVERSITY
SCHOOL OF ELECTRICAL ENGINEERING

ABSTRACT OF THE
MASTER'S THESIS

Author: Kaisu Ölander

Title: Eye movements and early magnetoencephalographic brain responses induced by faces in natural scenes

Date: 30.10.2015 Language: English Number of pages: 7+67

Department of Neuroscience and Biomedical engineering

Professorship: Systems neuroscience Code: SCI018Z

Supervisor: Prof. Riitta Hari

Advisor: D.Sc. Linda Henriksson

Faces are a crucial part of social interaction and they affect our visual perception. In this thesis, I studied how the presence of a face in natural visual stimuli affects the eye movements, and which visual features in the images best explain the scanpaths and brain responses. A novel approach based on representational similarity was applied to relate the brain responses to the eye tracking results.

In the experiment, the subjects were shown grayscale photographs of social occasions. A part of these images were shown both upright and upside-down. The subjects' eye movements and magnetoencephalographic (MEG) brain responses were measured during the experiment. MEG responses were analyzed at sensor level. Representational similarity analysis (RSA) was applied in determining whether the low-level features or the faces in the images can explain the brain responses before the first saccade after the stimulus onset and if they can explain the eye movements during the viewing of the images.

The presence of a face had a clear effect on the eye movements, especially when the images were shown upright: the latency of the first saccade was 60 ms smaller, and the scanpaths were more similar between different subjects. The map depicting the locations of the face in the images explained the scanpaths better than did the traditional saliency maps.

The MEG responses at around 55–130 ms after stimulus onset were best explained by the saliency maps, reflecting the low-level visual features of the stimuli. However, at latencies 90–125 ms the MEG responses correlated also with the face maps and, importantly, at latencies 105–140 ms the responses correlated with the forthcoming scanpath.

Keywords: magnetoencephalography (MEG), natural scenes, face detection, representational similarity analysis (RSA), eye movements

AALTO-YLIOPISTO
SÄHKÖTEKNIIKAN KORKEAKOULU

DIPLOMITYÖN
TIIVISTELMÄ

Tekijä: Kaisu Ölander

Työn nimi: Kasvojen aiheuttamat silmänliikkeet ja varhaiset aivovasteet

Päivämäärä: 30.10.2015 Kieli: Englanti Sivumäärä: 7+67

Neurotieteen ja lääketieteellisen tekniikan laitos

Professuuri: Systeeminen neurotiede Koodi: SCI018Z

Valvoja: Prof. Riitta Hari

Ohjaaja: TkT Linda Henriksson

Kasvot ovat tärkeä osa sosiaalista kanssakäymistä, joten ne vaikuttavat visuaaliseen havainnointiin. Tässä diplomityössä tutkittiin luonnollisissa kuvaärsykkeissä esiintyvien kasvojen vaikutusta silmänliikkeisiin sekä sitä, mitkä kuvan piirteet selittävät parhaiten silmänliikkeitä ja aivovasteita. Uutta samankaltaisuuden arviointiin perustuvaa analyysimenetelmää käytettiin etsittäessä yhteyttä aivovasteiden ja silmänliikemittausten tulosten välillä.

Koeasetelmassa koehenkilöt katselivat harmaasävyisiä valokuvia sosiaalisista tilanteista; osa kuvista näytettiin sekä oikein- että väärinpäin. Koehenkilöiden silmänliikkeet sekä aivovasteet mitattiin kokeen aikana. Aivovasteiden mittaamisessa käytettiin koko pään kattavaa magnetoenkefalografiaa (MEG). Vasteita tutkittiin kanavatasolla. Samankaltaisuusanalyysiä sovellettiin selvittämään, selittävätkö kuvien matalan tason piirteet tai kasvojen paikat aivovasteita ennen ensimmäistä sakkadia tai silmänliikkeitä katselun aikana.

Kasvojen silmänliikkeisiin aiheuttamat vaikutukset olivat selkeät etenkin kuvissa, jotka näytettiin oikeinpäin. Näissä kuvissa ensimmäisen sakkadin latenssi oli 60 ms lyhyempi ja lisäksi silmänliikkeet muistuttivat enemmän toisiaan eri koehenkilöiden välillä. Kasvojen sijaintia kuvaava kartat selittivät silmänliikkeet paremmin kuin perinteiset salienssikartat.

MEG-vasteet 55–130 ms kuvan näyttämisajankohdan jälkeen korreloivat eniten kuvien matalan tason piirteitä kuvaavien salienssikarttojen kanssa. Latensseilla 90–125 ms MEG-vasteet korreloivat kuitenkin myös kasvokarttojen kanssa ja aikavälillä 105–140 ms myös tulevien silmänliikkeiden kanssa.

Avainsanat: magnetoenkefalografia (MEG), luonnolliset kuvat, kasvojen havaitseminen, samankaltaisuusanalyysi, silmänliikkeet

Preface

This thesis was done at the Department of Neuroscience and Biomedical Engineering, at the Human Systems Neuroscience group. It was funded by the Louis-Jeantet prize awarded to Riitta Hari. I want to thank my instructor Linda Henriksson for guidance and for giving me the possibility to work and learn in this interesting project. I am grateful to my supervisor Riitta Hari for giving me advice on this and possible future theses. I would like to thank also Mia Illman for assisting in the MEG measurements and Veli-Matti Saarinen for helping with the eye-tracking. Special thanks to Riku for huge amounts of support and making the writing of this thesis more cheerful.

Otaniemi, 30.10.2015

Kaisu Ölander

Contents

Abbreviations

EOG	electro-oculography
FDR	false discovery rate
fMRI	functional magnetic resonance imaging
MEG	magnetoencephalygraphy
SQUID	superconducting quantum interference device
RDM	representational dissimilarity matrix
ROI	region of interest
RSA	representational similarity analysis
SSS	signal-space separation
tSSS	spatio-temporal signal-space separation

1 Introduction

Faces give us essential information in everyday social interaction. By looking at faces, individual people can be distinguished from others and it is possible to tell whether they are in a good or bad mood. In addition, the importance of faces has been shown in multiple brain-imaging and eye-tracking studies. First, a region in the fusiform gyrus, the "fusiform face area", responds more strongly to the viewing of faces than other object stimuli [1]. Second, faces attract attention although the subjects would have been told to concentrate on something else [2, 3]. Furthermore, there is evidence that people tend to move their gaze faster towards an image if it depicts a face than if it does not [2]. Several eye-tracking studies have demonstrated that, during natural scene perception, the subjects are likely to look at faces [4] and they might have trouble looking away from a face [5].

Despite the popularity of face detection, there is still a shortage of studies which use stimuli depicting faces as a part of a natural scene. Particularly, few magnetoencephalography (MEG) studies have utilized this kind of stimuli and, moreover, there is a lack of studies in which the relationship between the MEG responses and eye movements in natural scenes is analyzed. Reasons for the shortage of such studies include the challenges to analyze the responses evoked by natural scenes and the contamination of the MEG signal by large saccades.

Several eye-tracking studies have employed saliency maps or equivalent models of stimulus images in predicting the eye movements occuring during viewing of natural scenes. However, comparing the brain responses with these kind of computational models has been challenging as the correspondency between the data from the brain and the model can not be defined straightforwardly. Representational similarity analysis (RSA) is a relatively new method enabling these comparisons. Nevertheless, while several functional magnetic resonance imaging (fMRI) studies have successfully used this method, few MEG studies have utilized RSA. Representational similarity analysis can reveal the brain-activity patterns related to certain experimental conditions better than univariate analysis methods [6], and, as MEG has a millisecond-range temporal resolution, RSA can be advantageous in resolving the

time course of information processing in the brain when applied in the analysis of MEG data.

The aim of this thesis is to determine how early the brain processes the information related to image features that affect the direction of the gaze. We presented grayscale photographs depicting natural scenes so that the images contained different amounts of faces, or no faces at all. A part of these images were shown both upright and upside-down to determine how the inversion of the image affects the processing of the image features and the gaze patterns. Both brain responses and eye movements were measured while the subjects were freely viewing these images. The brain signals were measured utilizing magnetoencephalography, and an eye tracker was used to determine the timing and location towards which the subjects direct their gaze. The MEG responses were studied at sensor level and the analysis was restricted to the moment before the subject made a saccade so that the MEG artifacts caused by saccades dis not affect the results.

Eye movements were analyzed separately for different image categories: images containing a face, images with no face, upright images and inverted images. Additionally, feature maps depicting the low-level features and the locations of the faces in the images were constructed. Representational similarity analysis was applied to determine which feature maps best explain the eye movements and the points in time during which the feature maps correlate with the brain responses.

This thesis is divided into five chapters. Chapter 2 presents the background of this thesis, the eye movements, face-specific brain responses and measurement methodology. Chapter 3 explains the experimental setups and methods used in the data analysis. Chapter 4 presents the results of the experiments; Chapter 5 discusses the main results of this thesis.

2 Background

In this thesis, the early brain responses and eye movements during viewing of natural scenes are studied by employing magnetoenchephalography (MEG) and eye tracking. Section 2.1 explains the eye movements and the eye tracking methods, and Section 2.2 presents the origin of the MEG signal and the measurement methods. Sections 2.3 and 2.4 discuss gaze control and face perception more in detail.

2.1 Tracking of eye movements

Various eye-tracking studies aim to discover which parts of a scene the people tend to look at. This kind of information can be utilized in many fields of research, ranging from marketing and usability research [7, 8, 9] to studying physiological and psychological processes. The eye-movement features, e.g., the durations and locations of the fixations, have been widely studied in the context of reading [10]. Other research interests include, e.g., the visual search [11, 12] and scene perception [13, 14, 15]. Furthermore, eye tracking has been utilized in revealing abnormalities in eye movements and viewing patterns related to mental disorders [16, 17].

Several studies employ both eye tracking and brain imaging techinques. In MEG studies, the electro-oculogram is often utilized to detect the eye movements as they cause magnetic disturbances which contaminate the MEG signal [18, 19]. However, the MEG and eye tracking data have been combined while studying, e.g., the cortical activity preceding saccades [20, 21, 22], mental disorders [23, 24] and gaze-related brain processing [25].

In this thesis, eye movements were tracked from subjects while viewing images that depicted natural scenes. Attention was paid particularly to the first saccade after the stimulus onset, but also fixations and blinks during the viewing were analyzed. Section 2.1.1 explains the different types of eye movements and Section 2.1.2 presents eye tracking methods.

2.1.1 Eye movements

Eye movements are frequent as they are needed both in maintaining the sharp vision at the focus of the gaze and in relocating the focus [10, 26]. To see a region in detail, the fovea of the eye must be directed towards it, as it has the highest vision acquity. The moments when the eyes are relatively stable and focused on a particular stationary point are called fixations. Other important types of eye movements are the saccades, smooth-pursuit movements and vergence movements. Smooth-pursuit movements occur when eyes lock to a moving target, whereas vergence movements are caused by a target moving towards or away from a viewer [27].

The details of a scene can only be seen during a fixation [10, 28]. Although the eye is relatively stable at the point of fixation, it does miniature eye movements called tremor, drift and microsaccades [29, 30]. These miniature eye movements are essential to our vision because an image which is stabilized on the retina disappears from the sight [26].

Saccades are rapid, ballistic eye movements which redirect the fovea from one place to another about three times in a second [31, 32, 33]. A relationship exists between visual attention and saccade programming [34, 35, 36], but the visual perception is reduced during a saccade [28]. The direction and the distance, also known as the amplitude, of the saccade are predefined. After a saccade has been initiated, it can not be stopped and its trajectory can not be influenced [31, 32, 37]. However, it is possible to prepare another saccade before an ongoing saccade has been terminated.

2.1.2 Eye-tracking methods

In this thesis two different systems (SMI RED500 and EyeLink 1000 Long Range Mount) were used in eye tracking. Both of these systems are based on dark pupil eye tracking. In the MEG experiment, also vertical and horizontal electro-oculograms were measured.

The location of the subject's gaze is relevant information in many studies. To find the gaze position, one needs to be able to detect the eye movements and,

furthermore, separate them from the head movements [38]. In both eye trackers used in this thesis, the eye movements are detected by tracking the pupil and the corneal reflection (CR), as two ocular features are needed to be able to compensate for the head movements [39, 40].

The pupil is distinguished from the surrounding iris because it reflects the light differently. The infrared (IR) light makes this phenomenon even more effective as the pupil absorbs almost all of the IR light making it appear much darker than the iris [41]. Thereby it is possible to locate the pupil center. The corneal reflection, on the other hand, appears as a highlight of the eye. The brightest reflection comes from the front surface of the cornea [41]. By measuring the location of the CR in relation to the pupil center, the eye and head movements can be separated since the relative location stays fixed with minor head movements and changes with eye rotation [38].

Electro-oculography (EOG) is an eye tracking method based on measuring the electrical potential from the skin around the eye [41, 42]. This potential is due to the opposite polarities at the cornea and the retina. The corneo-retinal potential causes a steady electrical potential which changes as the eye moves. This change in the potential can be measured by placing a pair of electrodes at the opposite sides the eye. The weakness of EOG is that the movements of the eyelid also influence the electro-oculograms, and therefore the interpretation of especially the vertical eye movements is complicated [18].

2.2 Magnetoencephalography

Magnetoencephalography (MEG) is a non-invasive brain imaging method which measures the weak magnetic fields caused by brain activity [43]. It is a counterpart for the electoencephalography (EEG), as EEG measures the electric potentials caused by the same source activity. Both MEG and EEG have a temporal resolution of millisecond time scale, which is better than in any other non-invasive neuroimaging technique. In addition, the spatial resolution of MEG is better than that of EEG as the skull and scalp do not distort the MEG signal as much as the EEG signal [44].

2.2.1 Neural basis of MEG signals

Information processing in the brain is based on the neurons, which communicate by sending electrical impulses to each other [43, 45]. Information is passed on from one neuron to another at a synapse, the current arriving from the first neuron causes a postsynaptic current in the second neuron. The postsynaptic activity creates a primary current, which gives rise to passive ohmic currents, the volume currents, both within the cell and in the surrounding medium.

The magnetic fields caused by the primary and volume currents are the source of the MEG signal [43]. These currents are small, so a group of neurons must be activated in synchrony to produce a measurable magnetic signal. Furthermore, the direction and location of the electrical current have an impact on the MEG signal as well. The head can be thought as a nearly spherical conductor. In an ideal sphere, only currents that have a tangential component to the skull create a MEG signal and the radial currents do not produce any magnetic field outside the head. Consequently the currents in the fissures of the cortex account for the most of the MEG signals. The MEG signal is typically considered to be formed by the synchronous postsynaptic currents in the pyramidal cells that are located in the cerebral cortex.

2.2.2 MEG measurement

The magnitude of the magnetic fields caused by neural activity is typically 50-500 fT [43, 46]. The earth's geomagnetic field is multiple times stronger and even the human heart produces a stronger magnetic field. Currently, the only way to measure these small signals is to use the superconducting quantum interference devices (SQUIDs). The SQUID sensors are submerged in liquid helium, which is at a temperature of 4 K (–269 Celsius). The magnetic flux is transfered to the SQUID by flux transformers. There are multiple kinds of flux transformers, e.g., a single loop, magnetometer, and two oppositely wound loops, gradiometer. The gradiometers can be either axial gradiometers, in which the loops are placed vertically one above the other, or planar gradiometers, in which the loops are aligned in the same horizontal plane. The planar gradiometers can be called 'near-sighted' as they are sensitive to

the currents directly underneath it, whereas the magnetometers are most sensitive to the currents a few centimeters distance from the loop.

As the magnetic fields are extremely small, it is highly important to cancel the obtrusive magnetic fields [43]. Therefore the MEG recordings are performed in a magnetic shielded room, which minimizes the noise caused by electric or moving metallic objects, such as the nearby traffic and powerlines. However, it is impossible to block all the interferences outside the shielded room, but there are ways to separate the noise from the actual brain signal. A well-known and effective method for this is the signal-space separation (SSS) [47, 48]. In this method, the Maxwell's equations are used to get rid of the interfering signals in the MEG data. The data are divided into three subspaces: the brain signals which originate from inside the sensor helmet, interfering signals originating from outside the helmet and noise located close to the sensors. The subspaces are independent and the signal from outside the helmet can be removed.

One of the greatest challenges in the MEG studies is the electromagnetic inverse problem, which means the localization of the source currents. The inverse problem has no unique solution; it is not possible to deduce the cerebral current distribution based on the knowledge about the magnetic field measured from outside the head [43]. However, using physiological constraints and other information, physiologically plausible solutions are obtained. In this thesis, the sites of the source currents were not identified as the brain responses were studied only at sensor level, and thus also no anatomical MRI scans were acquired in this study.

2.3 Eye movements while viewing natural scenes

Section 2.3.1 presents theories regarding the gaze control and eye movements during scene viewing. Section 2.3.2 discusses the effects that a presence of a face in a scene has in the eye movements.

2.3.1 Gaze control

During scene perception, people direct their gaze to informative and interesting regions of a scene [15, 49]. These regions are fixated more often, and the fixation durations are longer [13, 50, 51, 52]. However, there are no unambiguous explanations describing the properties which make a region interesting. For instance, the viewing task affects the eye movements [53, 54, 55]. Two models have been proposed to explain how the gaze control works during scene perception: top-down knowledge-driven gaze control and bottom-up stimulus-based gaze control.

According to the top-down control of attention, the gaze is controlled by cognitive and visual systems. Knowledge from previously seen similar scenes and short-term memory during the current scene are utilized in eye movement control [15]. It is known that a model including contextual information predicts human fixations better than the models with no top-down information [56], and the consistency of a target object within a scene affects the fixation patterns [50, 51, 52].

The bottom-up control of attention is based on the assumption that information from local image properties directs the viewer's attention. Uniform regions in a scene are not considered interesting. Instead, the points of fixation have often different spatial features, for example a higher contrast and lower correlation with surrounding pixels [57], than the neighbouring regions [15]. Attempts have been made to explain the points of fixation with saliency maps which topographically indicate the visual saliency of an image. For example, in the well known saliency map model by Itti and Koch [58, 59], visual saliency is topographically computed based on color, intensity and orientations in the image. These low-level feature maps are combined to form one saliency map in which the stimulus saliency has been shown to correlate with fixation locations in complex natural scenes [60].

2.3.2 Eye movements induced by faces

Although saliency maps based on low-level features have successfully predicted the points of fixations in natural scenes, they can not explain the strong bias towards faces. For example, Birmingham et al. [61] found that the saliency maps by Itti and Koch [59] performed poorly with social scenes. On the other hand, Cerf et al. [62] showed that adding a semantic feature channel, which includes e.g. faces and text, to the saliency map based on color, intensity and orientation improved the prediction of fixations.

Moreover, a face strongly attracts attention regardless of the task. Crouzet et al. [2] reported that people tend to look at the face even though they would have been instructed to look at something else and Gilchrist et al. [5] showed that, if a face is present, people have difficulties looking away. Additionally, Bindemann et al. [63] showed that an upright face as a distractor stimulus delayed the response times in a go/no-go task more than other distractor stimuli. On the other hand, Crouzet et al. [2] discovered that the latencies of saccades triggered by faces are shorter than those triggered by something else, such as a vehicle.

2.4 Face-specific respones in the human brain

Faces used as visual stimuli induce fast saccades and attract attention more than other stimuli. An interesting question therefore concerns the timing of recognizing a stimulus representing a face as a face. Numerous MEG studies e.g. [64, 65, 66, 67] have found face-sensitive deflections, M100 and M170, peaking around 100 ms and 170 ms after stimulus onset respectively. These responses can be caused by other stimuli as well, but are stronger for natural face stimuli. Additionally, EEG studies have shown that visual face stimuli produce a chain of event-related potentials (ERP), P100 and N170, at similar latencies e.g. [68]. Both functional magnetic resoncance imaging (fMRI) and MEG studies have also located the source of the face-specific responses to the the fusiform gyrus [64, 65], which is located in the inferior temporal cortex [1, 69, 70].

Some MEG studies suggest that the M100 response is related to the low-level features [65, 67], whereas the M170 response is more sensitive to higher level perception [66, 67]. Similar results have been reported by EEG studies; by using intact and phase-scrambled images of cars and faces as stimuli, Rossion et al. [68] concluded that the face-sensitivity of P100 and N170 is driven by different features of the stimulus. In addition, Crouzet et al. [2] argued that while in their experiment the first saccades towards the face were triggered within 100 ms, the face during this time had not been completely recognized as a face. Therefore, the saccades towards the face were probably initiated due to low-level cues.

In this thesis the faces were a part of a natural scene, as opposed to the previous studies, in which the faces have been extracted from the natural background. Additionally, the subjects were not advised to look at the faces, but to view the images freely. After the viewing of each image the subjects had to answer a simple yes/no -question about the image, the questions were not directly related to specific objects in the image. The aim in this thesis was to determine how the faces, as a part of a natural scene, affect the eye movements and brain responses and, furthermore, how the analysis of these two can be combined.

3 Materials and methods

This study comprised two parts. First an eye-tracking only study was conducted to validate the stimulus set and to test the experimental setup. The second experiment included MEG measurements and was conducted after the analysis of the eye tracking results from the first experiment. This section describes the stimulus set, the experimental setup and the stages of data analysis.

3.1 Stimulus set

The stimulus set consisted of 199 grayscale photographs depicting natural scenes. The images had different topics, for example wedding, athletics, dinner, or work. The images were divided into 4 categories depending on whether they contained one face, two faces, three or more faces, or no face. All images were collected from Wikimedia Commons™ and they were chosen so that each topic comprised images from all categories. Figure 1 shows example images.

3.1.1 Image preprocessing

The original images in Wikimedia Commons™ were mainly colour images and the dimensions of the images differed greatly. All 199 images were cropped to the size of 1400 x 1050 pixels by using Adobe® Photoshop® in such a way that faces in the original images were located diversely in all parts of the cropped images; see Appendix A for details. The images were converted to grayscale by using MATLAB® R2014a. In addition, the intensity values of these grayscale images were restricted from the range 0–1 to the range 0.2–0.8 to get smoother results from the luminance matching of all images. The matching was done with a function provided by the SHINE toolbox [71] so that the desired mean (M) and the desired standard deviation (S) for the matched images was computed as the average value of all means and standard deviations of the original images. The new luminance-matched images I_{new} were obtained by

(a) Image with no face.

(b) Image with a face.

(c) Image with two faces.

(d) Image with three (or more) faces.

Figure 1: Example images from each image category: images with no face, one face, two faces, and three or more faces. The fixation cross was visible during the viewing. After each image, the subjects had to answer a simple yes/no -question about the image, e.g. "Could you see water in the image?"

$$I_{new} = \frac{I_{orig} - m}{s} S + M, \tag{1}$$

in which m is the mean and s is the standard deviation of the original image I_{orig}. The SHINE toolbox provides also the possibility to match the luminance histograms of the images. However, this option was not used as the original histograms differed significantly from each other and too much noise would have been introduced to the resulting images.

3.2 Experimental setup

The setups were similar in both experiments. The aim of the eye-tracking-only experiment was to validate the stimulus set and experimental setup. Based on the results of this experiment, some minor changes were made in the setup for the experiment including MEG recording.

3.2.1 Gaze experiment

Nineteen subjects with normal or corrected to normal vision participated in the experiment. The subjects were from 21 to 29 years old, mean age was 24.7 years. Twelve of the subjects were female and seven male. The experimental setup was programmed in Python, using the Psychopy framework [72]. The eye tracker used in the experiment was an SMI RED500. The gaze data were sampled at 500 Hz. The eye tracker was attached to the bottom of the monitor, which was 60 cm away from the subject. The size of the image on the screen was 41.5 cm x 30 cm, so that viewing angle was approximately 38 degrees. During the experiment the subject's head leaned against a chin-forehead rest. The experiment was conducted in a dimly lit room. The 9-point calibration was done only once in the beginning of the experiment.

During the experiment, the 199 images were shown in random order in four sets. From each set, 10 images were randomly chosen and shown both upright and upside-down. Thus, each set contained 60 images, except the last one, which contained 59 images. The inverted images were shown among the original images and the order was randomized.

The images of a set were shown for either 0.5 s, 1.0 s, 1.5 s or 3.0 s. The order of the durations was balanced across subjects. Before each image the subject was shown a fixation cross for a time randomly chosen from between 0.8 and 1.6 s. This decreased the effects of the subjects' anticipation for the upcoming image. The fixation cross was visible over the image but subjects were advised to view the image freely. After each image the subject had to answer a simple question about the image, e.g. "Did the image contain a boat?". By clicking the left button of the mouse, the subject

answered "yes", and by clicking the right button, the subject answered "no". The question was shown for 5.0 s and during that time the subject had to answer. It was not possible to change the answer afterwards.

3.2.2 Combined gaze and MEG experiment

Eighteen subjects with normal vision participated in the experiment. The subjects were between 18 and 39 years old, mean age was 26.9 years. Ten of them were male and eight female. The experiments were conducted at the MEG core of Aalto NeuroImaging in Otaniemi with a 306-channel MEG device (Elekta Neuromag^{TM}, Elekta Oy, Helsinki, Finland) in a 3-layered magnetically shielded room (Imedco AG, Hägendorf, Switzerland).

Psychopy [72] was again used in building the experimental setup. The eye tracker used in this experiment was an SR Research EyeLink1000 Long Range Mount, with 500-Hz sampling rate. In this experiment the subject was sitting with the head inside the MEG helmet, and therefore no chin-forehead rest was used. The eye tracker was placed on a table in front of the subject. The monitor was 105 cm away from the subject and the size of the images were 72 cm x 54 cm and hence the viewing angle was approximately 38 degrees. The 9-point calibration was done twice during the experiment, once in the beginning and second the half way through the experiment. The room was dimly lit.

The same stimulus set was used in this experiment as in the eye tracking only part, but some minor changes were done in the setup. Each set contained 75 images, except for the last part which contained 74 images, 25 images were shown upright and 25 were shown both upright and upside-down. All of the images were shown for 1.0 s and the duration of the fixation cross was randomly chosen between 1.3 and 2.1 s. The questions were answered by lifting the left finger on the response pad for "yes" and lifting the right finger for "no".

3.3 Data analysis

The data consisted of eye-tracking data and MEG data. Section 3.3.1 introduces the saccade-detection algorithms used in converting the raw eye-tracking data. The similarities of the subjects' scanpaths were studied by employing the ScanMatch toolbox which is presented in Section 3.3.2. Representational similarity analysis (RSA) was applied in the MEG data analysis, see Section 3.3.4.

3.3.1 Saccade-detection algorithms

Two different eye trackers were used in this study, SMI RED500 and SR Research EyeLink1000 Long Range Mount, Section 2.1.2 for the working principles of these eye trackers. The raw data collected from the eye trackers were not used in the analysis, the fixations and saccades were detected by using the algorithms provided by the eye tracker manufacturers.

The raw data collected from the SMI RED500 eye tracker were stored in an .idf data file. The conversion from the raw eye tracking data to the fixations, saccades and blinks was done with IDF Event Detector 3.0.18. High-speed event detection was used with the following parameters: 40 °/s as the peak velocity threshold, 22 ms as the minimum saccade duration and 50 ms as the minimum fixation duration. The start of the peak velocity window was at 20% of the saccade length and the end at 80% of the saccade length.

In the high-speed event detection, the saccades are extracted as primary events from the raw data [39]. To find saccade-like events, all the velocities are computed from the raw data. A saccade-like event is not yet necessarily a saccade; it has to exceed a minimum saccade duration and the peak-velocity value of the saccade must lie within a peak velocity window. Furthermore, a saccade-like event could also be a blink. The saccade-like events were first searched by detecting the points in which the velocity was higher than a given peak velocity threshold. Second, the first velocity before the peak and after the peak, which are lower than a given threshold, were searched. The distance between these two velocities was computed, to check the duration of the saccade-like event. In a typical saccade, the velocity of the eye

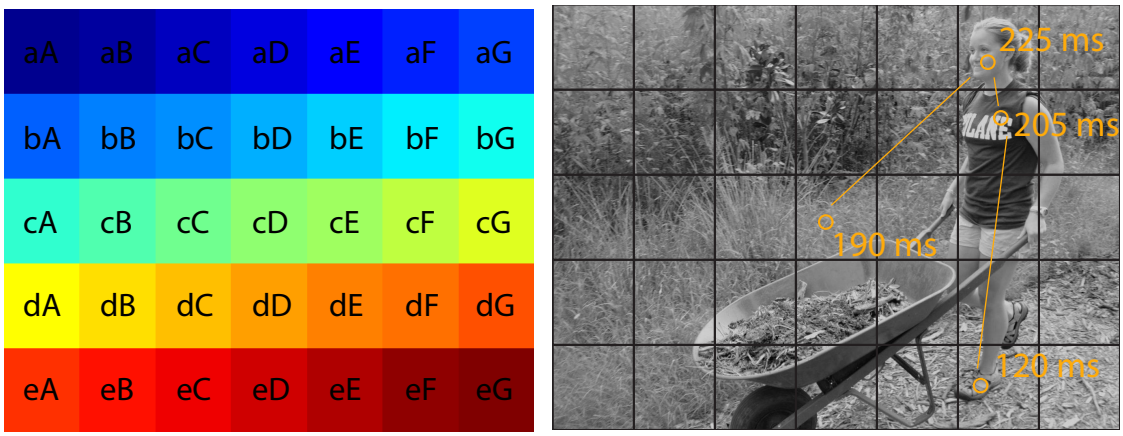
movements first rises, reaches a peak and then declines. The peak velocity must therefore lie within a certain peak-velocity window, which starts at a given point after the initiation of the saccade and ends at another point after the initiation of the saccade. These points are given as percentages of the saccade length (e.g. start is at 20% of the saccade length and the end at 80% of the saccade length). The speed of the pupil diameter change was checked in case of the blinks. If the speed exceeded an internal threshold, the saccade-like event was assumed to be a part of a blink. If the saccade-like event was not a blink and fulfilled the conditions given above, it was assumed to be a real saccade. The fixations were assumed to start after the previous saccade or blink and end before the next saccade or blink.

As for the SR Research EyeLink1000 Long Range Mount eye tracker, the raw eye tracking data was stored along with the information about the fixations, saccades and blinks in an .edf data file. As opposed to the SMI RED500, the EyeLink1000 eye tracker does the event detection by using an on-line parsing system. The velocity and acceleration of the eye movements were used in detecting the saccades. The .edf files were converted to ASCII format with EDF Converter 3.1.

3.3.2 Comparison of scan paths

The similarities of scan paths were analyzed by using the ScanMatch toolbox [73]. ScanMatch is based on the Needleman-Wunsch sequence alignment algorithm which has earlier been used in comparing DNA or protein sequences. In the ScanMatch toolbox this algorithm is used in quantitatively scoring two eye-movement sequences. The method takes into account both the relationship between the region of interest (ROI) and the fixation durations.

To compute the score depicting the similarity of two sequences, the image first needs to be divided into ROIs; see Figure 2 for details. This scoring can be achieved by dividing the image into regular bins or by using some designated areas of the image as ROIs. A letter combination is then given to each region, as shown in Figure 2. The fixations get tagged, in a chronological order, with the letter combination according to their location. To take into account the fixation duration, the tag of a fixation was repeated in the sequence proportionally to the fixation duration.



(a) The image divided into 5 x 7 regions of interest (ROIs) with unique letter combinations [73].

(b) An example image with four fixations and fixation durations. The fixation at the time of stimulus onset is in the center of the image.

Figure 2: The letter combinations of each region in the image are shown in (a). The fixations shown in (b) are converted to a sequence by tagging them chronologically with the letter combination of their location. By applying 50-ms bins the sequence becomes: cDcDcDaFaFaFaFbFbFbFbFeFeF.

To find the best alignment, the Needleman-Wunsch sequence alignment algorithm uses local optimal alignment of sub-sequences [73]. Optimization is accomplished by scoring all possible alignments. The algorithm has two parts: creating a matrix with all possible scores and finding the alignment with the highest score. In the first part, the matrix is created by using a substitution matrix which contains information about the relationship between ROIs. In this study, this relationship was based on the inverse Euclidean distance between the regions, meaning that the lowest score was given to the bins furthest apart and highest to the bins at the same location. A threshold can be chosen to determine at which point the values of the substitution matrix become negative. As the substitution matrix and the length of the sequences affect the magnitude of the score, the scores need to be normalized by

$$s_n = \frac{s_{orig}}{max(\mathbf{M_s}) * l}, \tag{2}$$

where s_n is the normalized score, s_{orig} the original score, $max(\mathbf{M_s})$ the maximum value of the substitution matrix and l the length of the longest sequence. Figure 3 shows an example of scoring two sequences. The algorithm allows gaps to be added

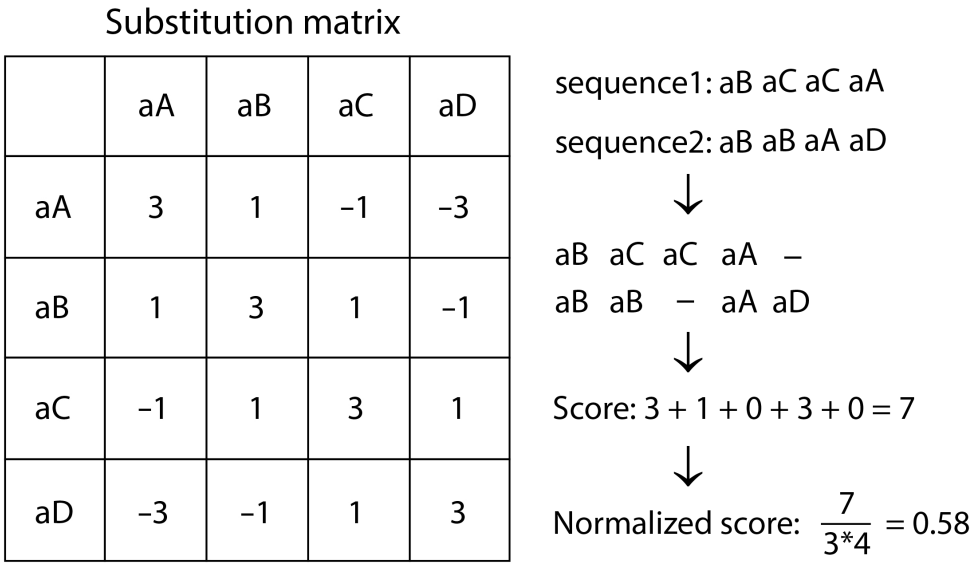


Figure 3: A simple example of a substitution matrix and scoring of two sequences. First the two sequences are aligned to give the best possible score between the sequences. No gap penalties were used in this example. Finally the score is normalized.

to the sequences and it is possible to set the value of a gap penalty. A negative gap penalty encourages the alignment of distantly related regions and constricts the insertion of gaps.

In this experiment the ROIs were created by dividing the image to 11 x 9 bins. Thus, the letter combinations on the first line of the grid were between aA – aK and on the last line between iA – iK. Another possibility could have been to use the face regions in the images as ROIs but this would have caused a problem when comparing the scan paths in the images which contained no face. Temporal binning of 50 ms and a threshold of 3.5 was used.

3.3.3 Saliency and face maps

The visual saliency in the images was detemined by computing the Itti and Koch saliency maps [58, 59], as well as face maps depicting the locations of the faces in the images. The standard Itti and Koch saliency map is based on the local color, intensity, and orientation information in the image. However, as our images were grayscale, the computation of the maps was based on intensity and orientation.

In the Itti & Koch model, the image regions with local spatial discontinuities are considered to be visually salient. The discontinuities are detected by applying a center–surround organization, which is a common structure in the receptive fields of the retinal ganglion cells [74, 75]. This organization resembles two concentric circles, the center (a smaller circle) and the surround (a larger circle). The receptive fields are sensitive to differences between the visual features in the center and the surround, e.g., a bright spot in a dark background.

To discover the spatial discountinuities, raw feature maps were first extracted from the image, in this case the feature maps comprised the intensity and orientation maps. This was accomplished with Gaussian pyramids, which low-pass filter and subsample the image [76]. The original image was low-pass filtered and subsampled to produce a new image which was decreased in size and resolution. The process was repeated and as a result a set of images, which resembled a pyramid, was obtained.

The orientation and intensity, which are early visual features, can be extracted from Gaussian pyramids. A Gaussian pyramid $I(\sigma), \sigma \in [0, .., 8]$, is obtained from the intensity image I, which is the $I(0)$ in the pyramid. The center–surround operation is implemented as the difference between two images at different levels in the Gaussian pyramid. The images at levels $c \in \{2, 3, 4\}$ represent the center and the images at levels $s = c + \delta, \delta \in \{3, 4\}$ represent the surrounding. The images representing the surrounding were interpolated to the finer scale and the difference was computed by point-by-point subtraction. This procedure resulted in six different feature maps for intensity.

The local orientation differences between the center and surrounding were computed by applying a Gabor filter at four orientations to the images on all different levels in the pyramid. After the center–surround operation, 24 orientation feature maps $O(c, s, \theta)$ were obtained, in which $\theta \in [0°, 45°, 90°, 135°]$.

The six intensity feature maps and 24 orientation maps were normalized to obtain two conspicuity maps: intensity (I) and orientation (O). The maps were normalized in such way that the strong distinguishable peaks in the maps were emphasized while the impact of homogeneous peaks was diminished. The normalized conspicuity maps were summed to obtain the final saliency map

$$S = \frac{1}{2}\left(\mathcal{N}(I) + \mathcal{N}(O)\right). \tag{3}$$

The face maps were computed by marking the center of each face in the images. Thereafter, the face maps were filtered with a Gaussian lowpass filter to create Gaussian blobs on the center coordinates of the faces to represent the faces in the images. The size and standard deviation of the Gaussian filter was determined by computing the average size of all the faces in the stimulus set. The size of the faces was computed as $d_i = \sqrt{h_i w_i}$, in which h_i is the height and w_i the width of the i:th face [77]. In addition, a map combining both saliency and face maps was constructed by computing the average of these two maps.

3.3.4 Representational similarity analysis (RSA)

Representational similarity analysis can reveal commonalities and dissimilarities between brain responses and computational models related to certain experimental conditions [78]. Different experimental conditions are associated with distinct brain-activity patterns [79]. Comparing these brain-activity patterns with behavioural data and computational models has been challenging. RSA makes these comparisons possible with the help of a representational dissimilarity matrix (RDM) which characterizes the similarities between the different brain-activity patterns, not the activity patterns themselves.

The brain-activity patterns can be seen as the representation of the stimulus, and the RDM depicts the distance between two representations [78]. This distance can be defined by, e.g., $1 - c$, where c is correlation, between the representations associated with two different stimuli. All representations are compared pairwise, thereby resulting in a dissimilarity matrix in which each cell represents the dissimilarity between two representations. A basic RDM is a square matrix in which the different stimuli are in the same order both vertically and horizontally, making the matrix symmetrical. The diagonal is zero as it depicts the dissimilarities between the same stimuli. The representations can also be derived from a model characterizing the stimulus. The similarities between the brain responses and the model can then be

examined by comparing the dissimilarity matrix constructed from the model and the RDM constructed from the brain-activity data. The representations can therefore be compared between, e.g., different subjects, brain-activity measurement techinques (MEG and fMRI), and brain and behavioural data.

In this thesis the dissimilarities between the brain-activity patterns related to different stimuli were compared by computing $1 - c$. The brain-activity patterns were defined as the response measured at a certain point in time from all the MEG channels, as the responses differed spatially slightly between subjects and no other channel group could be unambigously chosen. Therefore the comparisons were done separately at every point in time from, e.g., –100 ms to 400 ms with respect to stimulus onset, see Figure 4, resulting in a three-dimensional RDM matrix (number of images (N_{img}) x number of images (N_{img}) x number of time points (N_t)).

The RSA has been eariler used with MEG data by, e.g., Cichy et al. [80], who measured fMRI and MEG responses to 92 images. The MEG data were classified pairwise between all images at time points from 100 ms before to 1200 ms after the stimulus onset to create a 92 x 92 decoding matrix. Two different fMRI dissimilarity matrices were constructed by extracting the voxel values from the primary visual cortex (V1) and inferior temporal (IT) cortex. By correlating these two RDMs with the MEG decoding matrix, it was possible to show that the MEG signals correlated first with V1 and later with IT activity.

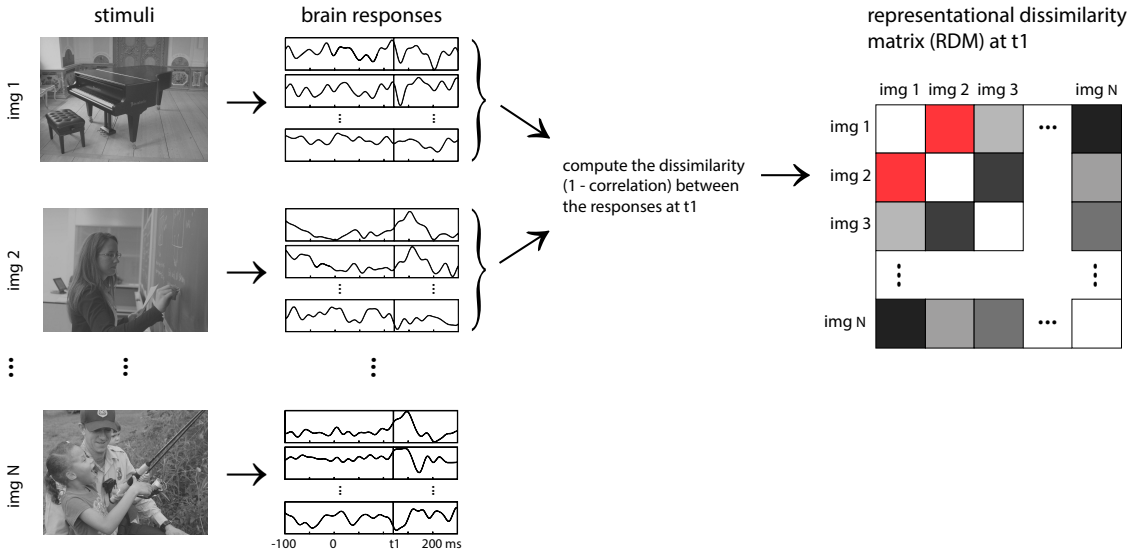


Figure 4: Schematic picture depicting the formation of the representational dissimilarity matrix (RDM). Different stimuli induce different brain-activity patterns and the dissimilarities between these patterns can be compared by e.g. $1 - c$, where c is correlation. In this example, the brain responses are measured from all different MEG channels, and the brain-activity pattern is defined as the MEG signal in all channels induced by a stimulus at a certain point in time (t1) with respect to stimulus onset. The dissimilarity matrix is therefore three-dimensional (N of stimuli x N of stimuli x N of time points). In this example, the stimuli are organised in the same order both vertically and horizontally. Therefore the matrix is symmetric and the diagonal is zero.

Instead of the fMRI data, in this thesis the RDMs were correlated at each point in time with scanpath similarity matrices to detect resemblances between the direction of the first saccade and the brain responses before the first saccade, Section 4.2.4 presents the results. Additionally, the RDMs were constructed by computing the distances between the brain responses induced by the upright and the inverted images. In this case the diagonal of the RDM representsed the (dis)similarity between the responses induced by the same image shown upright and upside-down. By comparing these values with the off-diagonal values it was possible to determine the points in time during which the same image shown upright and upside-down induced more similar responses than two different images; see Section 4.2.3.

4 Results

In this chapter, the results of both Gaze experiment and Combined gaze and MEG experiment are presented. The eye tracking data are analyzed in Section 4.1, with the aim of finding differences in the viewing patterns between images with a face and without a face. The results from the MEG data analysis are shown in Section 4.2. The analysis consists of examining resemblances between the responses induced by upright and inverted images, and correlations between the brain responses and eye movements. Both the eye tracking and MEG data were analyzed with MATLAB® R2014a.

4.1 Eye-tracking results

The eye-tracking data was analyzed separately from Gaze experiment and Combined gaze and MEG experiment. It seems that substantially the results were similar. The eye tracking data were filtered down to saccades, fixations and blinks with methods described in Section 3.3.1, the raw eye tracking data were not used in the analysis. The analysis focused on comparing the gaze patterns between the different categories: images containing a face, images not containing a face, upright images and inverted images. The gaze patterns were studied by analyzing the the latency and direction of the first saccade, the similarity of the scanpaths and the duration of the fixation of the first saccade.

Both experiments included some trials in which the saccade latency was less than 80 ms or in which the subject did not have time to make any saccades during the image. These trials, and trials in which the subject did not look at the fixation cross at the image onset, were excluded from the analysis (10% of all trials). The saccades with latencies less than 80 ms could be due to the subject failing to fixate on the fixation cross properly as the amplitude of these saccades was often quite small.

The data in Gaze experiment were collected from 19 subjects. The data from different image durations were analyzed separately. The data in Combined gaze and MEG experiment were collected from 18 subjects. New subjects were recruited for

Combined gaze and MEG experiment, although one of the subjects had participated in the piloting of Gaze experiment. In this experiment all the images were shown for 1.0 s and therefore all data were analyzed together.

4.1.1 The latency of the first saccade was shorter for images containing faces

A point of interest is whether the presence of a face in an image had an effect on the latency of the first saccade after the stimulus onset. The latencies of the first saccade were studied separately for upright and inverted images to determine if the inversion of the image affected the latency.

In Gaze experiment, the latencies in the upright images containing a face were 63.4 ± 6.9 ms shorter than in the images with no face, see the Figure 5. The trials in which there was a face in the image, but the saccade did not land on the face, were also included. The difference between the average latencies in images containing a face and with no face was tested with a paired, two-sided Wilcoxon signed rank test. At all image durations, if the image was shown upright, the latencies (mean ± SEM) were statistically significantly shorter for images containing faces than for images with no faces: at the image duration of 0.5 s the latencies were (242.2 ± 5.6 ms vs. 300.0 ± 8.4 ms; $p = 0.0001$), at 1.0 s (251.3 ± 5.5 ms vs. 324.8 ± 13.5 ms; $p = 0.0003$), at 1.5 s (259.4 ± 9.1 ms vs. 312.3 ± 9.7 ms; $p = 0.001$), and at 3.0 s (266.6 ± 5.9 ms vs. 336.6 ± 20.7 ms; $p = 0.004$).

The latencies of the first saccades in Gaze experiment were studied separately in images which were shown upside down, and the difference between the average latencies in images containing a face and with no face was tested with a two-sided Wilcoxon rank sum test. The latencies (mean ± SEM) in the images containing a face were distinctly shorter only in the cases in which the images were shown for 1.0 s (281.6 ± 12.8 ms vs. 335.9 ± 26.9 ms; $p = 0.02$) or 1.5 s (284.2 ± 9.8 ms vs. 339.5 ± 24.7 ms; $p = 0.04$), see the Figure 5. At image duration of 0.5 s (267.4 ± 7.6 ms vs. 287.1 ± 18.9 ms; $p = 0.2$) and 3.0 s (308.5 ± 12.5 ms vs. 294.6 ± 20.1 ms; $p = 0.5$) the latency differences were not statistically significant. In this experiment, only 10 images were shown both upright and upside down in all image durations.

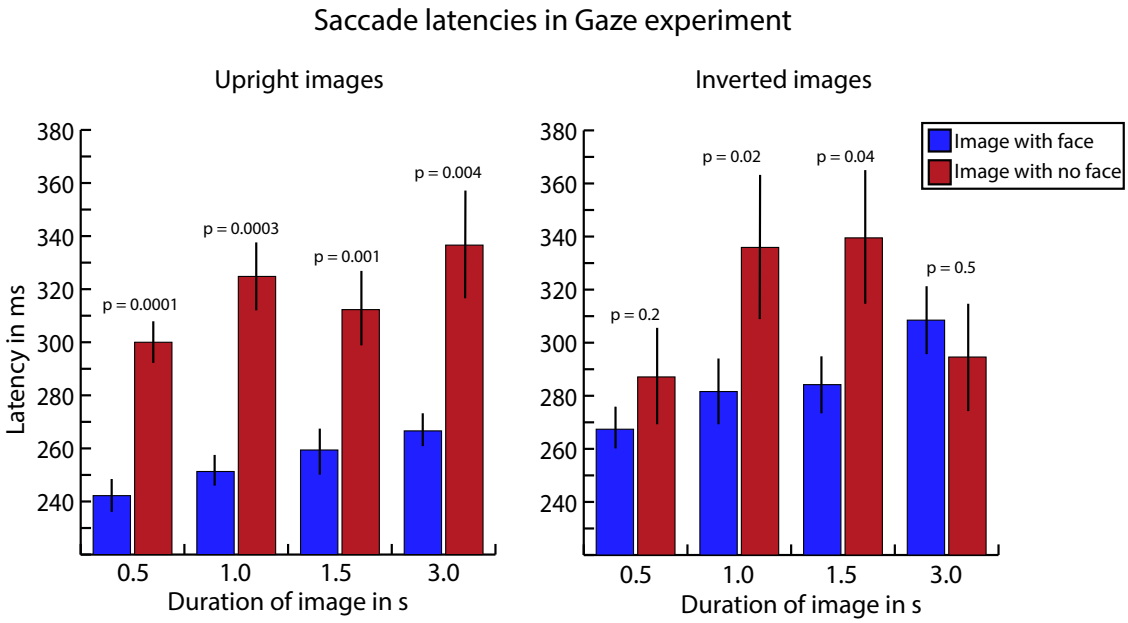


Figure 5: The (mean ± SEM) latencies of the first saccades across all subjects in the upright and inverted images, Gaze experiment. The latencies were shorter for images containing a face at all image durations if the image was shown upright. In the inverted images, the latencies of the first saccades were not unambiguously shorter if a face was present.

Therefore, in this case the difference between the face and no face conditions was not as evident, but this could be due to the insufficient amount of data.

Additionally, the difference between the average saccade latencies in the upright and inverted images was computed from the data gathered from Gaze experiment. Not surprisingly, the latencies were larger for the inverted images. The difference (mean ± SEM) across all subjects was at image duration of 0.5 s 26.7 ± 9.8 ms, at duration of 1.0 s 24.9 ± 10.8 ms, 1.5 s 24.2 ± 7.8 ms, and 3.0 s 21.8 ± 13.0 ms. The order of whether the upright or inverted image was shown first did not have a significant influence on the saccade latencies.

The difference between the saccade latencies in images containing faces and with no faces was repeated in Combined gaze and MEG experiment, as is shown in the Figure 6. If the image was shown upright, the first saccade was initiated 60.1 ± 4.6 ms earlier if the image contained a face. The difference between the average latencies in images containing a face and with no face was tested with a paired, two-sided

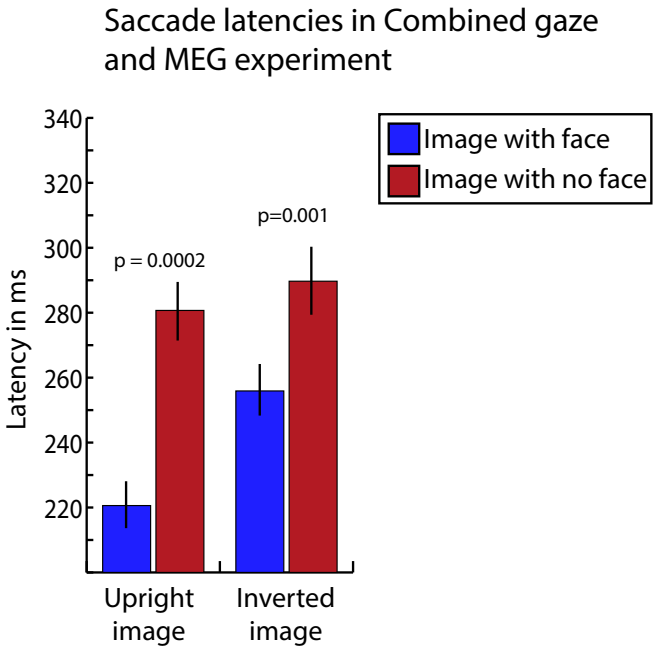


Figure 6: The (mean ± SEM) latencies of the first saccades across all subjects in the upright and inverted images, Combined gaze and MEG experiment. The latencies were shorter in images containing a face despite the orientation of the image.

Wilcoxon signed rank test. In this experiment, 100 images were shown both upright and upside down for each subject. Possibly due to the larger amount of data, the latency difference was statistically significant also in the inverted images. In upright images, the latencies (mean ± SEM) in images containing faces and images with no faces were (220.6 ± 6.9 ms vs. 280.7 ± 9.2 ms; $p = 0.0002$), and in inverted images (255.9 ± 8.4 ms vs. 289.7 ± 10.1 ms; $p = 0.001$).

In Combined gaze and MEG experiment, the difference (mean ± SEM) between the latencies in the upright and inverted images across all subjects was 31.4 ± 4.3 ms. If there was a face in the image, this difference was higher (36.6 ± 4.1 ms; $p = 0.0002$) than if there was not (8.1 ± 8.8 ms; $p = 0.3$). This can also be seen in the Figure 6, the inversion of the image has a notable impact on the saccade latencies only if a face was present. The latencies between upright and inverted images not containing a face did not differ significantly from each other. The order of whether the upright or the inverted image was shown first did not have a significant influence on the latencies.

4.1.2 Faces prolong the fixation duration

The fixations after the first saccade after stimulus onset were studied in order to see whether the faces affect the fixation duration too. The ROIs defined in the Section 4.1.3 were used in determining whether the fixation was directed to a face.

According to the data from Gaze experiment, the duration (mean ± SEM) of fixations to a face were significantly longer if the duration of the image was 1.0 s (36.6 ± 4.1 ms; $p = 0.002$), 1.5 s (36.6 ± 4.1 ms; $p = 0.01$) or 3.0 s (36.6 ± 4.1 ms; $p = 0.0003$) than the durations of fixations to something else, see Figure 7. The difference between the average durations of fixations to a face and to a region without a face was tested with a paired, two-sided Wilcoxon signed rank test. However, there was no significant difference between the durations of fixation to a face and to a region without a face if the duration of the image was 0.5 seconds. This might be due to the short viewing time of the images. Both the upright and inverted images were included in this analysis.

The data from Combined gaze and MEG experiment were utilized to see if there was a difference in the fixation durations in the cases in which the image was shown upright and upside down. In the upright images the duration (mean ± SEM) of fixations to a face were statistically significantly longer than to a region without a face (180.6 ± 5.3 ms vs. 167.6 ± 6.2 ms; $p = 0.02$), see Figure 7. However, the difference was not statistically significant if the images were shown upside down ($p = 0.3$).

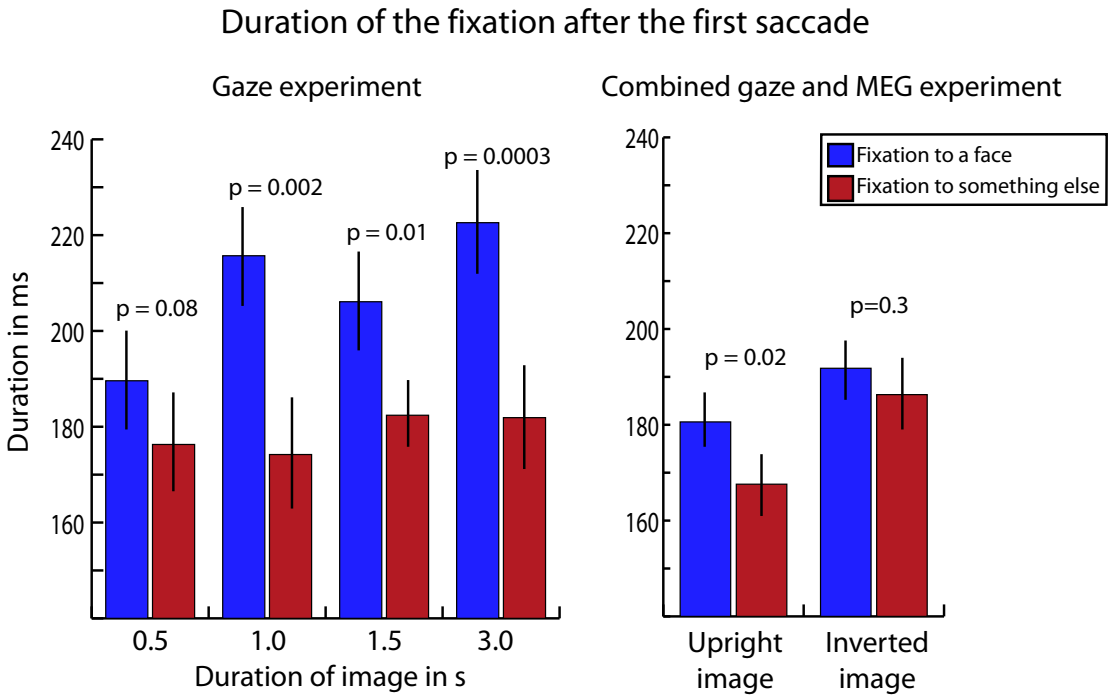


Figure 7: The (mean ± SEM) duration of the fixation after the first saccade across all subjects. In Gaze experiment, the durations were significantly longer to a face if the image was shown for 1.0 s, 1.5 s, or 3.0 s. In Combined gaze and MEG experiment, the durations of fixations to a face were significantly longer if the image was shown upright, but not if the image was shown upside-down.

4.1.3 A face in the image affects the direction of the first saccade

As numerous previous studies have shown, faces attract attention. To see whether already the first saccade can direct the gaze towards a face, the percentage of how often the first saccade lands on a face was studied. These percentages were computed separately for upright and inverted images by marking the face regions in the images. These regions were slightly bigger than the actual faces in the images in order to take into account the saccades which clearly were directed towards a face but landed just before the face.

Duration of image	Upright images	Inverted images	p-value
0.5 s	69.4%	45.6%	0.0005
1.0 s	67.2%	52.1%	0.0002
1.5 s	69.6%	53.8%	0.008
3.0 s	67.8%	56.5%	0.007

Table 1: The average percentages of how often the first saccade landed on a face in Gaze experiment. The percentages were computed across all subjects in images containing a face.

Duration of image	Upright images	Inverted images	p-value
1.0 s	67.5%	51.3%	0.0002

Table 2: The average percentages of how often the first saccade landed on a face in Combined gaze and MEG experiment. The percentages were computed across all subjects in images containing a face.

In both experiments, the probability that the first saccade lands on a face was quite high, as is shown in the Tables 1 and 2. In the upright images the average percentage was almost 70%. However, if the image was shown upside down, the probability of the saccade landing on a face was only around 50%. All differences between the percentages were statistically significant.

Another point of interest was to determine how often the first saccade lands on the same part of the image despite the orientation of the image, and whether the presence of a face in the image affected this. An instance of the first saccade landing on the same face in both cases, when the image was shown upright and upside down, is shown in the Figure 8.

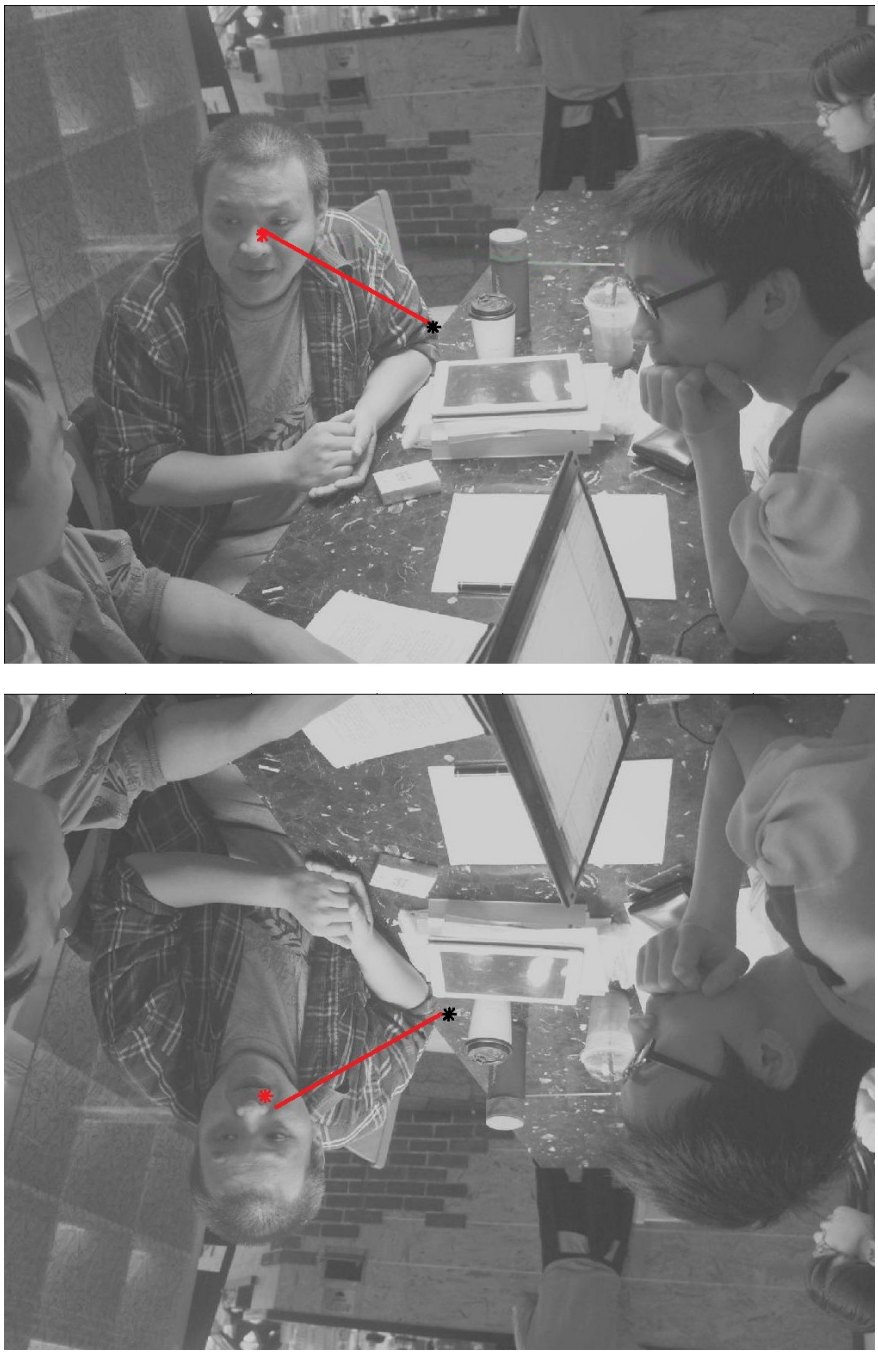


Figure 8: An example of the first saccade landing in the same part of the image despite the orientation of the image. The fixation in the onset of the image is shown as black dot, the first saccade and the fixation after it are shown in red.

The saccades were determined to land on the same part of the image if the distance between the end points of the saccades was less than 130 pixels. This distance was computed as the distance between the corresponding regions in the images

(see Figure 8). The size of the faces in the images varied considerably, making it challenging to determine a suitable distance. Therefore in some cases the saccades might have landed on the same face, yet in the analysis these are not classified as cases in which the saccades land on the same part due to the long distance between the end points. The distance of 130 pixels was applied as it is roughly in the same magnitude as the size of the ROIs (127x117 pixels) used in the scanpath similarity analysis, see Sections 3.3.2 and 4.1.4.

In Gaze experiment, the difference between the probabilities that the first saccades landed on the same part in the image, was significantly higher in images containing a face than in images without faces only if the duration of the image was 1.0 s ($58.1 \pm 3.5\%$ vs. $31.8 \pm 9.9\%$; $p = 0.01$), see Figure 9. The difference between the percentages in images containing a face and with no face was tested with a two-sided Wilcoxon rank sum test. The percentages (mean $\pm$ SEM) across all subjects of the saccades landing on the same part if a face was present were also higher if the images were shown for 1.5 s ($54.8 \pm 6.6\%$ vs. $38.8 \pm 9.9\%$; $p = 0.2$) or 3.0 s ($47.3 \pm 4.3\%$ vs. $39.2 \pm 10.7\%$; $p = 0.3$), but the differences were not statistically significant. This might be due to the insufficient amount of data, as few images were shown upside down in Gaze experiment.

Likewise, the percentages of the saccades landing on the same part of the image were higher if the image contained a face in Combined gaze and MEG experiment. In this case the percentages (mean $\pm$ SEM) across all subjects were computed separately for different amount of faces, see Figure 9. The probability of the saccades landing on the same part in the images in which there was no face ($44.4 \pm 3.2\%$) was statistically significantly lower than the probabilities of the saccades landing on the same part in images with one face ($60.2 \pm 3.8\%$; $p = 0.002$), and three or more faces ($57.8 \pm 3.6\%$; $p = 0.01$). The probability was higher also in images containing two faces ($52.5 \pm 4.1\%$; $p = 0.2$), but the difference to images with no faces was not statistically significant. The differences between the percentages of the saccades landing on the same part were not statistically significant between the images containing different amount (one, two, or three or more) faces.

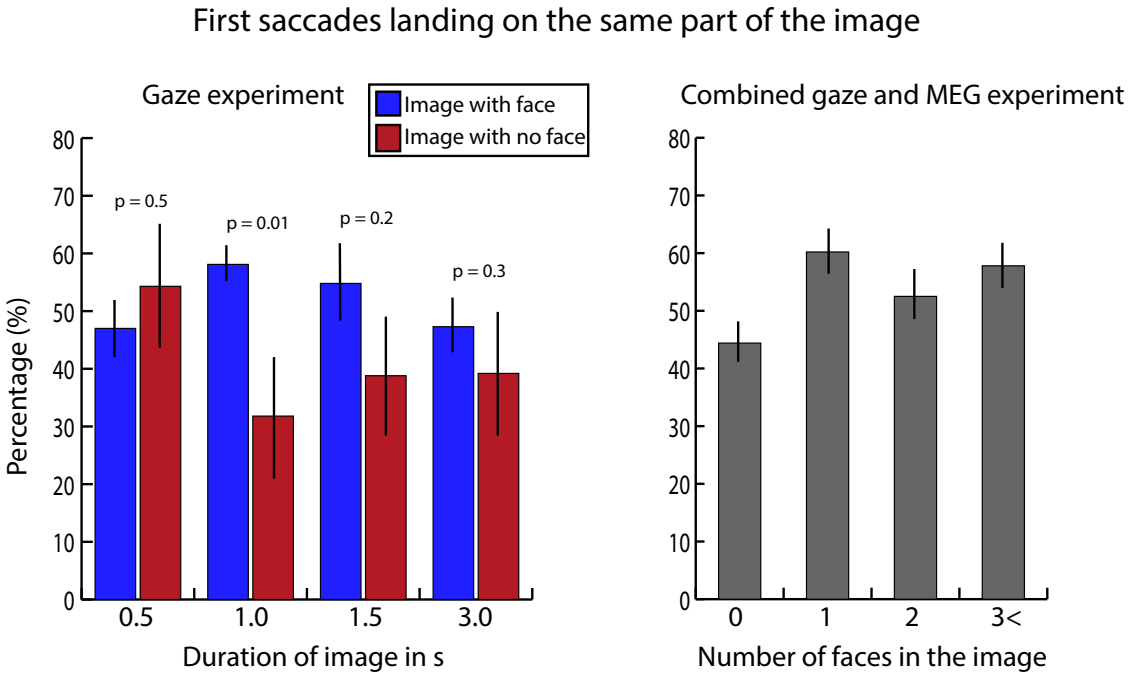


Figure 9: The (mean ± SEM) percentages of how often the first saccades landed on the same part of the image when it was shown both upright and upside down. In the Gaze experiment, the difference between percentages in images containing a face and images without a face was statistically significant only at the image duration of 1.0 s ($p = 0.01$). In Combined gaze and MEG experiment, the percentages were significantly higher in images with one, and three or more faces, than in images with no faces ($p <= 0.01$).

4.1.4 The similarity of scanpaths is higher between images containing a face than between images with no face

Whether the subjects had more similar viewing patterns in images containing a face than in images without a face, was studied by comparing the subjects' scanpaths in each image. The scanpaths were compared with the ScanMatch toolbox as described in the Section 3.3.2. This toolbox compares two scanpaths at the time and computes a score which depicts the similarity of these two scanpaths. The scores are normalized between 0 and 1, and the higher the score, the more similar the two scanpaths are. The images which were shown for 0.5 seconds in Gaze experiment were excluded from this analysis.

The averaged scores from both experiments show a significant difference ($p <= 0.002$)

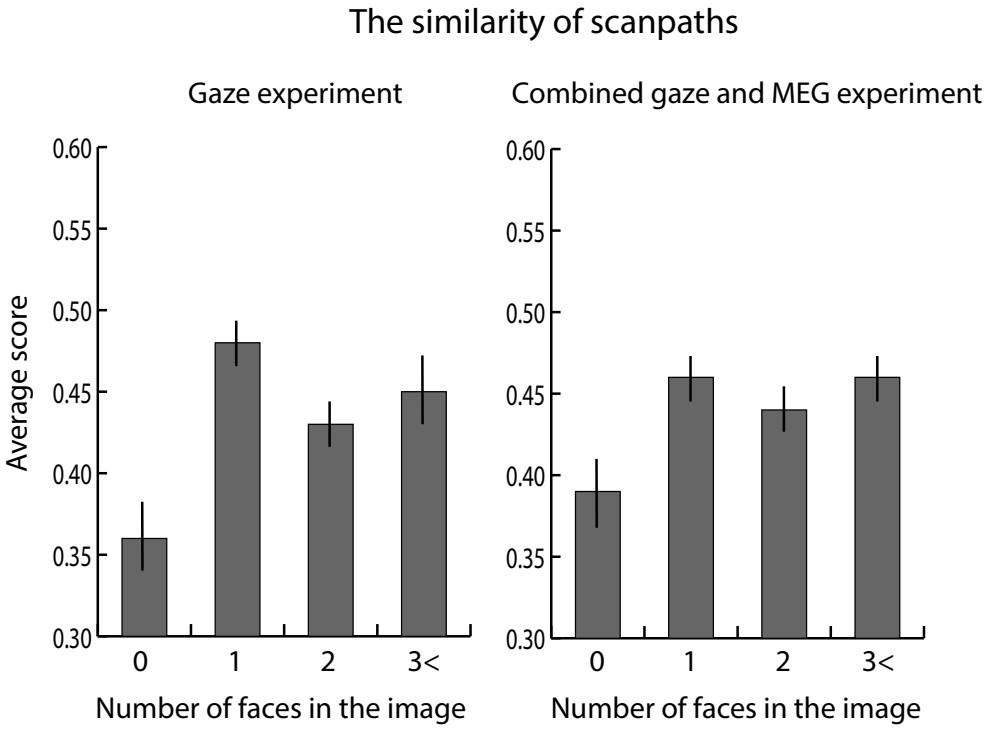


Figure 10: The (mean ± SEM) normalized score in images with different number of faces. In both experiments, the scanpath similarity scores were statistically significantly higher in images containing faces than in images without faces ($p <= 0.002$). The difference between scores in images with one face and two faces was also significant in Gaze experiment ($p = 0.005$), but not in Combined gaze and MEG experiment ($p = 0.2$).

between the similarities of the scanpaths in the images in which there is a face and in which there is not, see Figure 10. These scores were computed for all subject pairs in each image separately, and only the upright images were included in the analysis. In Gaze experiment the difference between similarities in images with one face and two faces was statistically significant as well ($p = 0.005$). This difference, however, was not repeated in Combined gaze and MEG experiment.

4.1.5 Face maps explain scanpaths better than saliency maps

To determine which features in the images affected the subjects' viewing patterns, the saliency and face maps were computed from the images. The saliency maps were computed by applying the Itti & Koch model, and the computation of the face maps was based on the locations of the faces, see Section 3.3.3. Three different maps depicting the image features were computed: a saliency map, a face map and a map combining both saliency and face information. The similarities of these maps were compared with the similarities of the scanpaths by applying the representational similiarity analysis (RSA), see Section 3.3.4.

Three representational dissimilarity matrices (RDM) of the visual saliency were construced by computing the distance ($1 - c$, where c is correlation) between the saliency maps, the face maps, and the combined saliency and face maps of each image pair. A fourth RDM was constructed for each subject separately by computing $1 - s$, where s is the scanpath score, between all image pairs. The scanpath scores from Combined gaze and MEG experiment were used in the analysis. The correlation between the scanpath score dissimilarity matrix and the RDMs of the visual saliency was computed separately for each subject.

The average correlation between all different RDMs of visual saliency (saliency maps, face maps and combined saliency and face maps) and the scanpath score dissimilarities across all subjects differed significantly from zero ($p = 0.0001$), see Figure 11. However, the correlation (mean $\pm$ SEM) between the RDMs of the face maps and of the scanpath scores was significantly higher than the correlation between the RDMs of the saliency maps and of the scanpath scores (0.29 ± 0.01 vs. 0.10 ± 0.008; $p = 0.0002$). This suggests that the faces explain the eye movements better than the saliency maps.

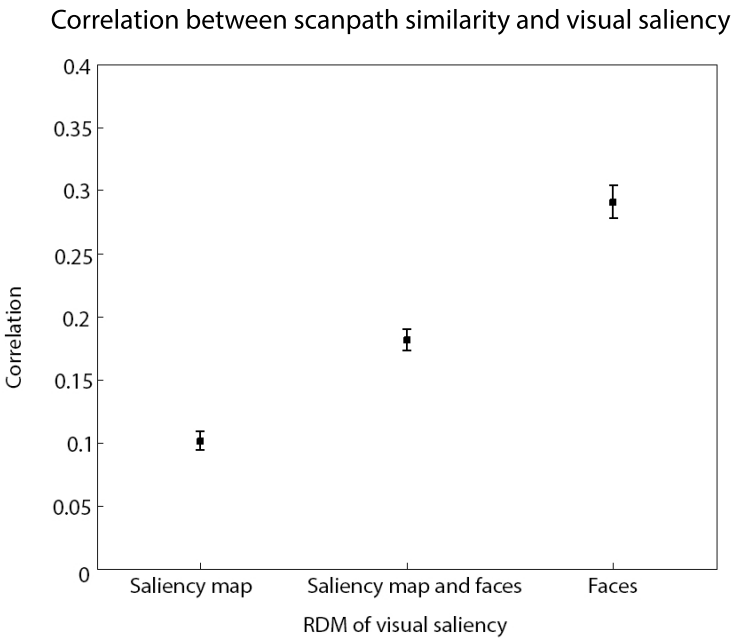


Figure 11: The average correlation between the RDMs of the visual saliency and scanpath score dissimilarities across all subjects in Combined gaze and MEG experiment. All correlations were significantly different from 0 ($p = 0.0001$) and all correlations differed significantly from each other ($p = 0.0002$). The (mean $\pm$ SEM) correlation between the RDMs of scanpath scores and facemaps was significantly higher than the correlations between the scanpath scores and saliency maps, and combined saliency and face maps.

4.2 MEG results

The MEG data were studied at sensor level. The sensor level responses differed notably between subjects and, furthermore, due to the complexity of the images, the responses in different image categories (e.g. the images containing a face and not containing a face) were not explicitly distinct. Therefore, the representational similarity analysis (RSA) was applied in the analysis. Some single-trial responses are represented in Section 4.2.2. The points in time during which the upright and the inverted images induced the most similar responses are determined in Section 4.2.3 and the correlation between the MEG signal and the direction of the first saccade is examined in Section 4.2.4.

4.2.1 Preprocessing of MEG data

The MEG data were preprocessed by using tSSS (MaxFilter) to suppress magnetic interferences. The correlation limit was set to 0.9 and the buffer length to 16 s. The bad channels were marked manually; the automatic detection of the bad channels while filtering the data was not used. The MaxFiltered MEG data were baseline corrected using the time window between 200 ms and 0 ms before the stimulus onset, and lowpass filtered at 45 Hz.

4.2.2 Single-trial and average MEG responses

The visual response were clearly visible in the MEG signals; Figure 12 shows average responses of one subject to images with faces measured from all gradiometers. However, the differences between the image categories could not be evidently observed in the single trial responses. Moreover, the sensor level responses differed notably between subjects. The responses induced by all images shown to one subject, measured from one magnetometer located in the posterior side of the MEG helmet, are presented in Figure 13. The responses were visible from at around 50 ms after the stimulus onset, but there was no systematic difference between the image categories.

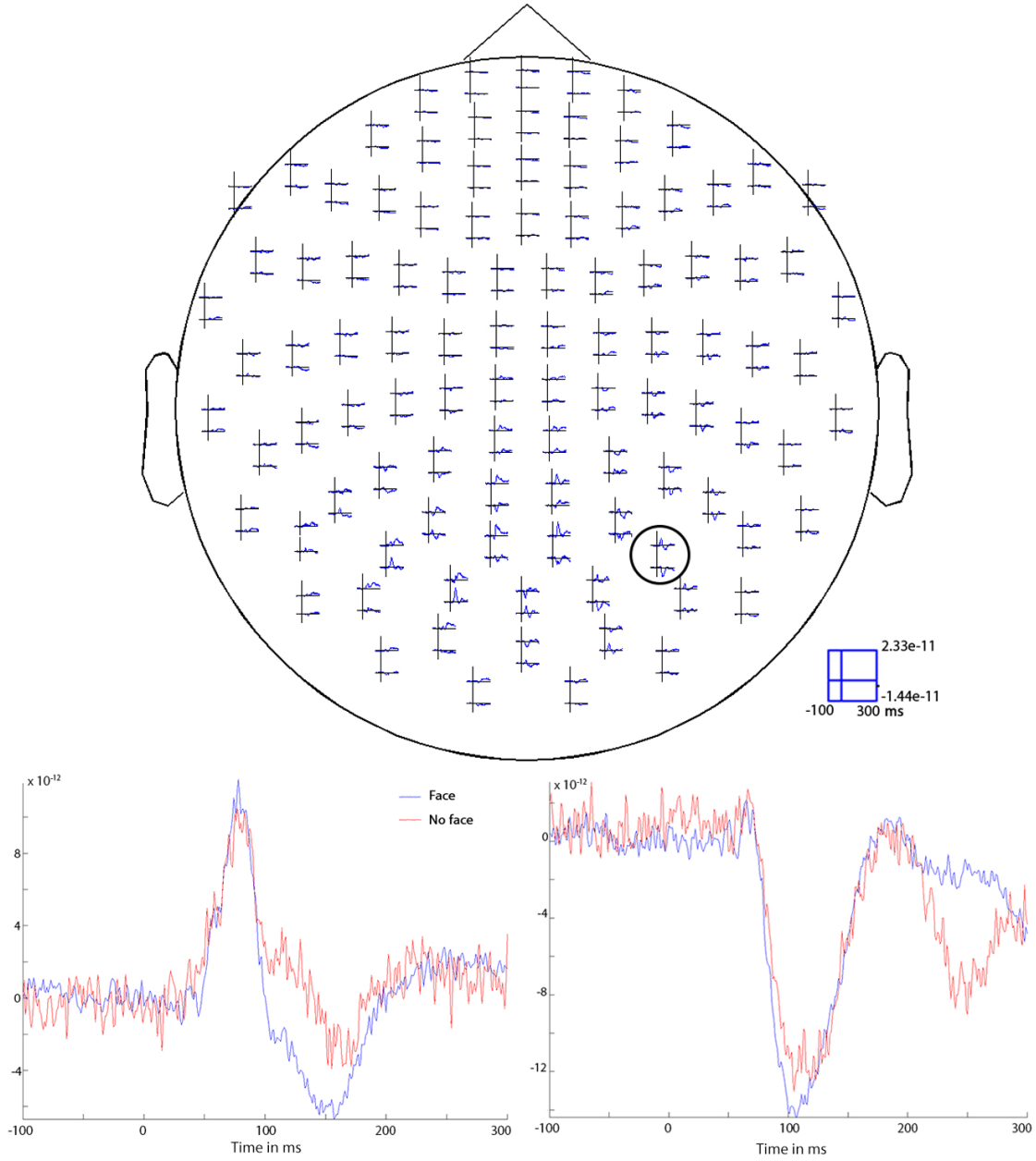


Figure 12: The image above represents the average responses of one subject to images with faces measured from the gradiometers, and the images below the whole head image represent the same subject's responses induced by images with faces and no faces measured from the gradiometer pair, the location of which is circled in the whole head image. The differences between the responses induced by these two image categories in this subject were most visible at this gradiometer pair.

The average responses induced by images with a face and with no face by two different subjects are represented in Figure 14. For both subjects, the responses induced by these two different image categories resembled each other, but the responses

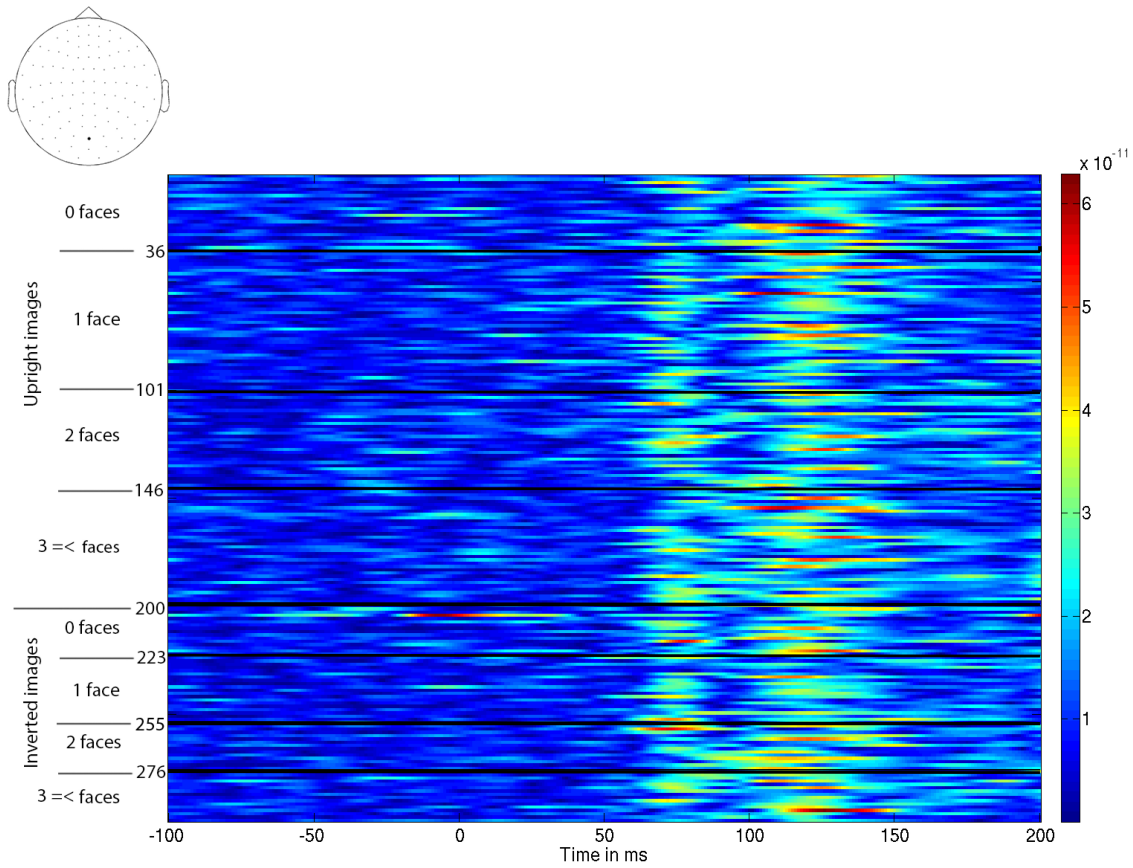


Figure 13: The single trial responses of one subject induced by the different image categories. The responses were measured as the square root of the power mean from one gradiometer pair, the location of which is shown in the image in the upper left corner. The response was clearly visible in all the images, but no distinct difference between the image categories could be observed.

differed substantially between the subjects. The responses were measured as the average from seven magnetometers or gradiometer pairs placed in three different locations, posterior left, middle and right side of the helmet, the exact location of the gradiometers is shown in the Figure 14. The responses from gradiometer pairs was computed as the square root of the power mean (see Equation 4).

$$M = \sqrt{\frac{1}{n}\sum_{i=1}^{n} {x_i}^2}, \quad n = 2 \tag{4}$$

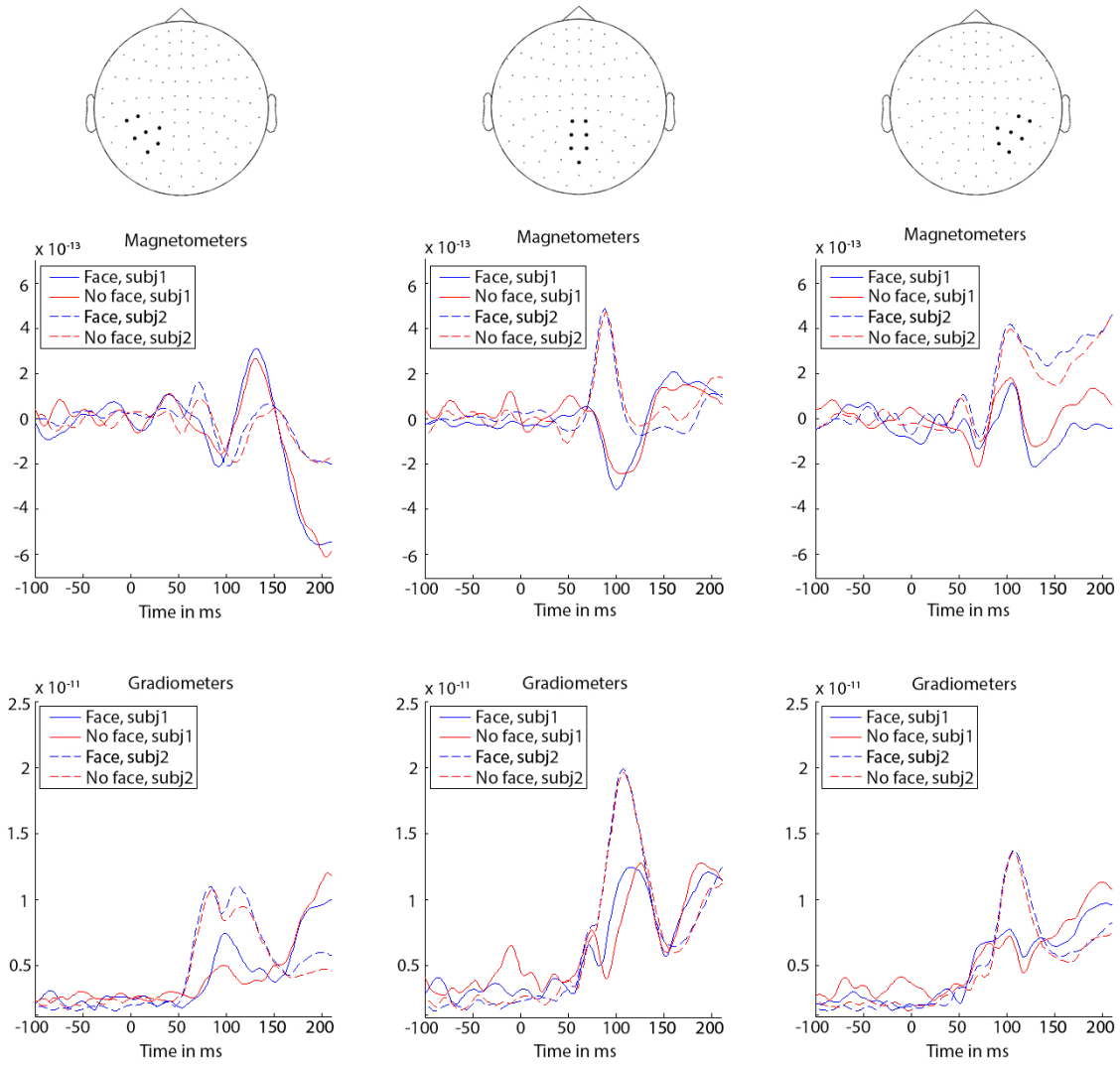


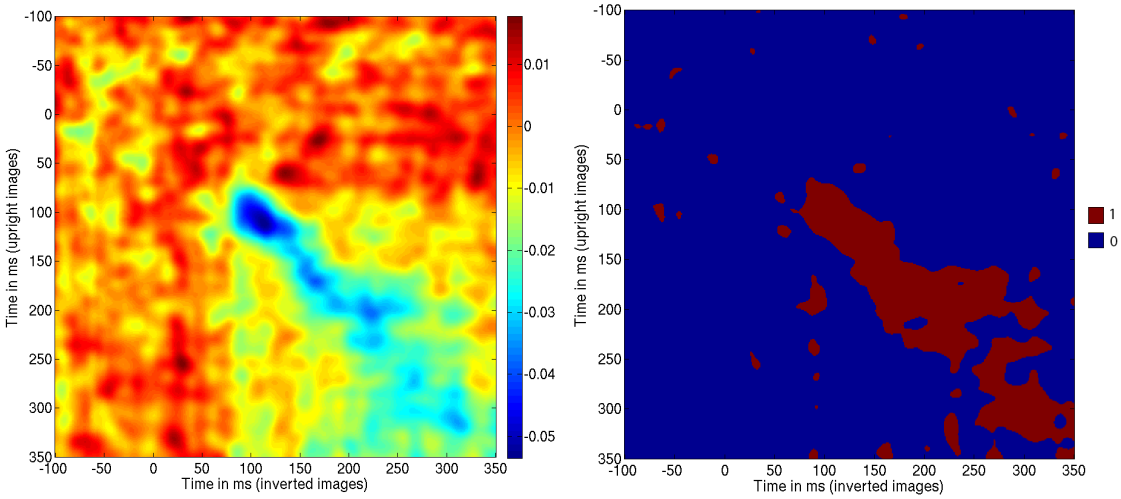
Figure 14: The brain respones induced by images containing faces and images not containing faces at three different locations from two subjects. The responses in the upper row were computed as the average responses from seven magnetometers located in the posterior left, middle and right side of the helmet. The responses in the bottom row were computed as the average responses from seven gradiometer pairs respectively. For both subjects, the responses induced by the two different image categories resembled each other, but the responses differed substantially between the subjects in both image categories.

4.2.3 Resemblances of responses induced by upright and inverted images

The aim of this analysis was to resolve at which points in time an image shown upside down produced more similar brain response with the same image shown upright, than the other images shown upside down. The purpose was to reveal whether the processing of an image is slower if it is shown upside down. The representational similarity analysis (RSA), see Section 3.3.4, was applied in the analysis.

The (dis)similarities of the brain responses were analyzed by constructing a representational dissimilarity matrix (RDM) containing the distances between the brain-activity patterns related to upright and inverted images. This distance was defined as the $1 - c$, where c is correlation, correlation between the brain responses induced by the images. The dissimilarity matrices were constructed separately for each subject, in the matrices the upright images were aligned vertically and the inverted images were aligned horizontally in the same order. Therefore the diagonal of the RDM represented the dissimilarities in the case in which the upright and the inverted images were the same picture and off-diagonal the dissimilarities in the case in which the images were different pictures. The images that were not shown both upright and upside down were excluded from the analysis.

In order to detect if the similarity between the brain responses was visible at a later point in time for the inverted image than for the upright image, the RDM was constructed from brain responses at different points in time for these images. This was accomplished by computing the dissimilarity between the brain response related to the upright image at time t_u and the brain response related to the inverted image at time t_i at all combinations of points in time (t_u,t_i), in which t_u and t_i are from 100 ms before the stimulus onset to 350 ms after the stimulus onset. The difference between the dissimilarities in the cases in which the pictures were the same and in which they were not, was computed by subtracting the off-diagonal values of the RDM from the diagonal values at all these points in time. The average difference can be seen in the Figure 15a.



(a) The similarity of the brain responses induced by the same picture shown upright and upside-down.

(b) The points in time during which the similarity was statistically significant at 5% significance level. False discovery rate (FDR) correction was used to correct for multiple comparisons.

Figure 15: The points in time during which the brain responses induced by the same picture shown upright and upside down were more similar than the brain responses induced by different pictures shown upright and upside down. The dissimilarity was computed as $1 - c$, where c is correlation, between the brain responses related to the upright and inverted images. The values in this matrix are the differences of the dissimilarities between the brain-activity patterns in the cases in which the upright and inverted images depicted the same picture and the case in which the pictures were different.

The same images induced a more similar response starting from at around 90 after stimulus onset, the responses were most similar at 112 ms after the stimulus onset for the upright images and 114 ms for the inverted images. After this point in time, the similar response is slightly (about 5 ms) delayed for the inverted images. This result is consistent with the single-trial responses which did not show a significant delay in the responses induced by the upside down images either, see Figure 13. The points in time during which the similarity was statistically significant at 5% significance level are shown in Figure 15b, false discovery rate (FDR) was used in correcting for multiple comparisons.

4.2.4 Correlations between brain responses and visual features

In the eyetracking data analysis the likeness of the eye movements induced by the images was studied by computing the similarity of the scanpaths, the results are presented in Section 4.1.4. Additionally, three feature maps (saliency maps, face maps and combined saliency and face maps) were computed from the images in order to determine which features best explained the scanpaths. The results, presented in Section 4.1.5, showed highest correlation between the scanpaths and face maps.

The aim of this analysis was to determine if the image features depicted in these maps accounted for the MEG data, and if there was a connection between the early brain responses and the upcoming scanpath. Representational similarity analysis (RSA) was applied in identifying the points in time in which the brain responses correlated with the feature maps and scanpaths, see Section 3.3.4 for a more detailed explanation of the analysis method.

Two different representational dissimilarity matrices (RDM) were constructed for each subject separately. The first RDM depicted the dissimilarity of the brain reponses induced by each image pair, the dissimilarity was defined as the $1-c$, where c is correlation, between the brain responses at a specific point in time. This RDM was first computed separately from the responses measured in the gradiometers and in the magnetometers. The average of the RDM of gradiometer responses and magnetometers was computed to form the final RDM of the brain responses. The dissimilarity matrix of the brain responses was computed at all points in time from 150 ms before the stimulus onset to 350 ms after the stimulus onset, though the correlations were studied only until the average saccade latency as after this the data might be contaminated by the eye movements. The second RDM depicted either the dissimilarities of one of the feature maps, or the dissimilarity of the scanpaths, between each image pair. The dissimilarities were computed as $1 - c$ between the feature maps and $1 - s$, where s is the scanpath score.

The highest correlation was found between the RDMs of the brain responses and saliency maps, see Figure 16. Surprisingly, the combined saliency and face maps did not explain the MEG data better than the original saliency maps. The correlation

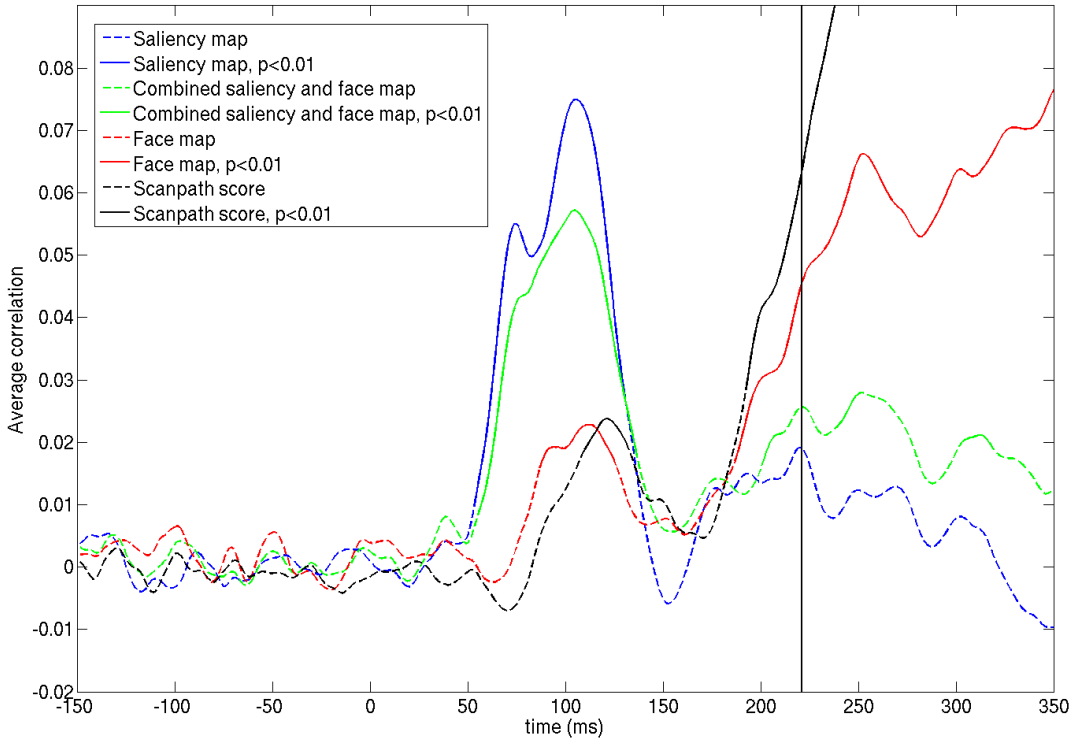


Figure 16: The correlation between the dissimilarities of the brain responses and the dissimilarities of feature maps and scanpath scores in upright images containing a face. The dashed lines represent the correlation across all subjects and the solid lines represent the points in time during which the correlation was statistically significant at 1% significance level. The black vertical line shows the average saccade latency.

between the MEG signal and traditional saliency maps was statistically significant from 55 ms to 130 ms after the stimulus onset, the first peak was at 75 ms. The highest peak in the correlations with the saliency maps and combined face and saliency maps was at 105 ms after the stimulus onset. The correlation between the RDMs of the brain responses and face maps was statistically significant from 85 ms to 125 ms, and peaked at 95 ms and 110 ms, after the stimulus onset. The correlation with the scanpath scores was statistically significant at latencies from 105 ms to 140 ms, and it peaked last, 120 ms after the stimulus onset. Only images with faces were included in this analysis, as it was not possible to construct face maps from images without faces. The correlations between the RDMs of the brain responses and scanpath similarities in the images without faces was noisy and no significant peaks were visible, suggesting that not enough data were available in order to complete the analysis.

The correlations were computed also by employing the brain responses induced by the images which were shown upside down, see Figure 17. In this case too, the saliency maps accounted best for the MEG data. The highest peak in the correlation between the RDMs of the brain responses and saliency maps was at 85 ms after the stimulus onset. The correlation between the brain responses and the combined saliency and face map peaked also at 85 ms after the stimulus onset, but the correlation between the brain responses and face maps was statistically significant only shortly from 80 ms to 95 ms after the stimulus onset, suggesting that the faces did not account as much for the MEG signals if the images were shown upside down. Nevertheless, in this case the correlations were noisier than with the upright images. Not as much data was available in this case, as only 100 images were shown upside down for each subject, which might affect the significance of the correlations. However, it might also be that the method used in computing the face maps was not ideal, as the saliency maps explained the MEG signal better than the face maps in the case in which the images were shown upright, too.

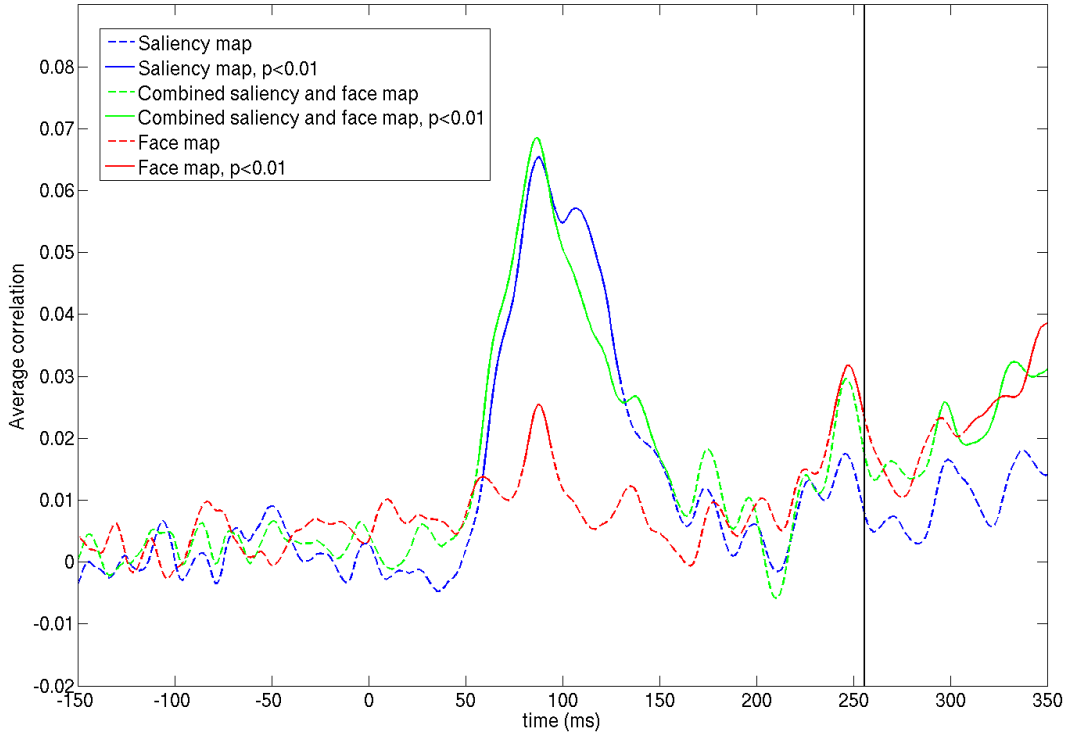


Figure 17: The correlation between the dissimilarities of the brain responses and the dissimilarities of feature maps and scanpath scores in inverted images containing a face. The dashed lines represent the correlation across all subjects and the solid lines represent the points in time during which the correlation was statistically significant at 1% significance level. The black vertical line shows the average saccade latency.

The face maps were created by computing Gaussian blobs to represent the faces. The center of the blob was located at the midpoint of the face it was representing. The blobs were computed by applying a Gaussian lowpass filter of size $[n_h, n_w]$ and with standard deviation σ. The size of the faces was computed as $d_i = \sqrt{h_i w_i}$, in which h_i is the height and w_i the width of the i:th face. The average size across all faces was $d_{avg} \approx 183$ pixels and standard deviation ≈ 96 pixels and, hence, these values ($n_h = 183$, $n_w = 183$ and $\sigma = 96$) were used in the Gaussian filter. However, as is shown in Figure 18, the filter parameters affected greatly the correlation between the MEG signal and the face maps. It seems a larger Gaussian filter might have given a higher correlation, but on the other hand, a negative correlation was introduced at around 70 ms after stimulus onset. Consequently, this might not have been the ideal method in constructing the face maps.

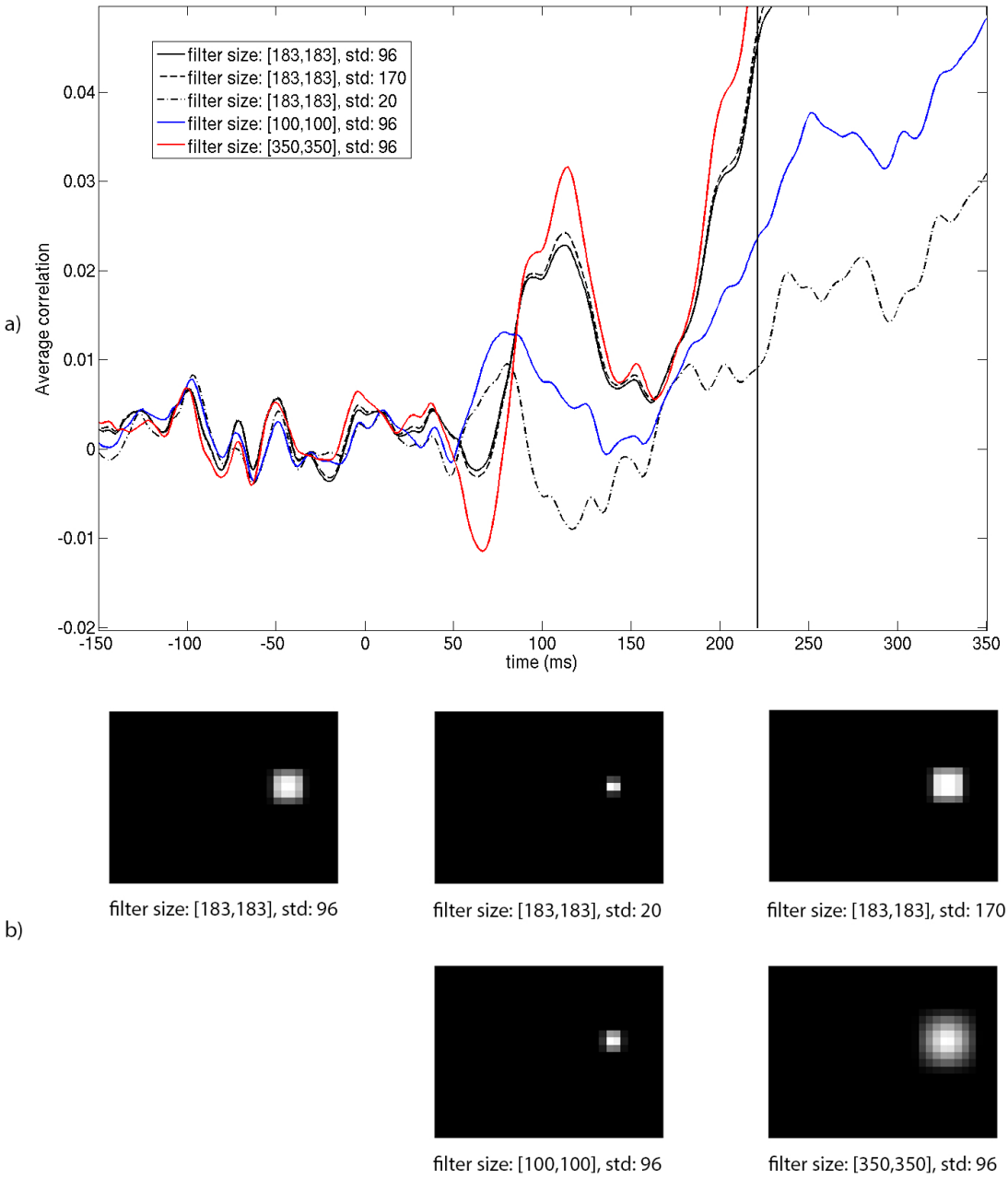


Figure 18: The effect of the filter parameters to the correlation between the MEG signal and face maps. (a) The average correlation across all subjects with different filter parameters. (b) The Gaussian blobs computed with the corresponding filters.

5 Discussion

This thesis aimed at characterizing the brain responses and eye movements induced by faces in natural stimuli and, furthermore, determining a method for relating the brain responses to the eye movements. One of the main results in this thesis was the significant difference between the latencies of the first saccade in images containing a face and with no face. If the image was shown upright, the first saccade was made approximately 60 ms earlier in images containing a face than in images without a face. Even in images which were shown upside down this difference was approximately 35 ms.

Generally the results in the eye tracking data analysis were similar in both Gaze experiment and Combined gaze and MEG experiment. Some figures, e.g., the saccade latencies and the fixation durations, differed between the experiments. This might be because different eye tracker models were used in the experiments. In addition, the conditions during the experiments were slightly different as in Combined gaze and MEG experiment the subject was sitting inside the MEG helmet. Furthermore, the duration of the images varied in Gaze experiment but not in Combined gaze and MEG experiment.

Fast saccades towards faces have earlier been reported by Crouzet et al. [2] who used a saccadic choice task in which the subjects had to target a face, an animal or a vehicle. They reported a significant difference between the mean saccadic reaction times (SRT) to faces (147 ms) and vehicles (188 ms). The fastest saccade latencies were just 100 ms, suggesting that the saccades were initiated before the face had completely been recognized as a face. However, in the study by Crouzet et al., the faces were mostly a main part of the images. In this thesis it was shown that the saccade latencies were shorter also when the faces were a smaller part of the scene. Additionally, in the study by Crouzet et al., the subjects were advised to saccade as quickly as possible to a target (e.g. a face or a vehicle), whereas in this study the subjects were allowed to view the images freely and the questions the subjects had to answer were not directly related to specific objects in the image.

In addition to the shorter latencies in the images with faces, the faces evidently affected the direction of the gaze. The first saccade landed on a face almost 70% of the time in both experiments, if the face was shown upright. Additionally, the scanpaths across all subjects were more similar and, furthermore, the first saccade was more likely to land on the same part of the image when it was shown upside down. In images containing two faces, the probability of the saccade landing on the same part was not significantly higher than in images containing no faces, but this could be due to the saccades landing on the other face when the image was shown upright as which they landed on when the image was shown upside down. The faces affected also the fixation durations; the average duration was longer in the individual fixations which were directed to a face, which is consistent with previous studies [81].

However, the effects the presence of a face has on the eye movements were not as evident in the inverted images. The latency of the first saccade was shorter in this case too if the image contained a face. On the other hand, the latency in the images containing a face was significantly shorter in the upright images than in the inverted images, and this difference was not found in images without faces. Additionally, in the inverted images, the probability of the first saccade landing on a face was only around 50%. Furthermore, the individual fixation durations to a face were significantly longer than the fixation to a region without a face only if the images were shown upright. On the other hand, the individual fixations to the faces were longer if the image was shown upside down, similar results have earlier been reported by, e.g., Barton et al. [82].

Three different feature maps were constructed to determine which features in the images best explained the scanpaths and the MEG responses. The feature map depicting the locations of the faces in the images explained the scanpaths significantly better than did the traditional saliency maps, or combined saliency and face maps. However, the correlation between the traditional saliency maps and the MEG responses was significantly higher than the correlation between the MEG responses and face maps. Furthermore, adding the face information to the saliency maps did not increase the correlation. This result leads to the conclusion that method used

in computing the face maps and combined face and saliency maps might not have been ideal.

In this thesis the face maps were created by computing Gaussian blobs in the locations of the faces. The blobs were computed by applying a Gaussian filter with a fixed filter size based on the average size of all faces in the stimulus set. In other studies, slightly different methods have been applied in constructing the face maps. These include constructing a binary heat-map by marking a minimally sized ROI around each face [62], and using different sizes for the filter based on the size of each individual face [4].

In addition, the computation of the combined face and saliency maps differed slightly from the method used by Cerf et al. [4, 62] who treated the face maps as an additional conspicuity map in the formation of the original saliency maps [58], therefore the saliency maps were computed as the average of four conspicuity maps (intensity, color, orientation and faces). In this thesis, however, the original saliency maps were computed first and the face information was added afterwards by computing the average of the saliency and face maps. It is possible that these methods could have explained the MEG responses better, but on the other hand, so far they have only been used to explain eye-tracking data, and the face map model in this thesis did explain the scanpaths better than the traditional saliency maps.

Nevertheless, all of the feature maps correlated statistically significantly with the MEG responses. The correlation between the MEG signal and traditional saliency map was significant at around 55–130 ms after the stimulus onset, whereas the correlation between the face maps was significant at 90–125 ms after the stimulus onset, and the correlation between the scanpath scores even later, 105–130 ms after the stimulus onset. The correlation between the traditional saliency maps returned to zero at around 150 ms after the stimulus onset, whereas the other correlations decreased but stayed slightly above zero. Additionally, the same image shown both upright and upside down induced most similar response at around 110 ms after the stimulus onset.

In summary, the MEG responses induced by the upright images correlated with the face maps and the upcoming scanpaths, although the saliency maps depicting the low

level information in the images explained the MEG responses better. Nevertheless, a face map was evidently the best model in explaining the scanpaths. Although from these data it is not possible to deduce the exact point in time at which it is decided where in the image the gaze will be directed, it can be assumed that some information related to this is processed at around from 105 to 130 ms after the stimulus onset.

References

[1] N. Kanwisher, J. McDermott, and M. M. Chun, "The fusiform face area: a module in human extrastriate cortex specialized for face perception," *The Journal of Neuroscience*, vol. 17, no. 11, pp. 4302–4311, 1997.

[2] S. Crouzet, H. Kirchner, and S. J. Thorpe, "Fast saccades towards faces: Face detection in just 100 ms," *Journal of Vision (Charlottesville, Va.)*, vol. 10, no. 4, pp. 1–17, 2010.

[3] M. L. Furey, T. Tanskanen, M. S. Beauchamp, S. Avikainen, K. Uutela, R. Hari, and J. V. Haxby, "Dissociation of face-selective cortical responses by attention," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 4, pp. 1065–1070, 2006.

[4] M. Cerf, J. Harel, W. Einhaeuser, and C. Koch, "Predicting human gaze using low-level saliency combined with face detection," in *Advances in Neural Information Processing Systems 20*, J. Platt, D. Koller, Y. Singer, and S. Roweis, Eds. Curran Associates, Inc., 2008, pp. 241–248.

[5] I. D. Gilchrist and H. Proske, "Anti-saccades away from faces: evidence for an influence of high-level visual processes on saccade programming," *Experimental Brain Research*, vol. 173, no. 4, pp. 708–712, 2006.

[6] R. D. Raizada and N. Kriegeskorte, "Pattern-information fmri: New questions which it opens up and challenges which face it," *International Journal of Imaging Systems and Technology*, vol. 20, no. 1, pp. 31–41, 2010.

[7] M. Wedel and R. Pieters, "A review of eye-tracking research in marketing," *Review of Marketing Research*, vol. 4, no. 2008, pp. 123–147, 2008.

[8] L. A. Granka, T. Joachims, and G. Gay, "Eye-tracking analysis of user behavior in www search," in *Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 2004, pp. 478–479.

[9] R. Jacob and K. S. Karn, "Eye tracking in human-computer interaction and usability research: Ready to deliver the promises," *Mind*, vol. 2, no. 3, p. 4, 2003.

[10] K. Rayner, "Eye movements in reading and information processing: 20 years of research." *Psychological Bulletin*, vol. 124, no. 3, p. 372, 1998.

[11] P. Viviani and R. G. Swensson, "Saccadic eye movements to peripherally discriminated visual targets." *Journal of Experimental Psychology: Human Perception and Performance*, vol. 8, no. 1, p. 113, 1982.

[12] J. M. Findlay, V. Brown, and I. D. Gilchrist, "Saccade target selection in visual search: The effect of information from the previous fixation," *Vision Research*, vol. 41, no. 1, pp. 87–95, 2001.

[13] J. R. Antes, "The time course of picture viewing." *Journal of Experimental Psychology*, vol. 103, no. 1, p. 62, 1974.

[14] G. R. Loftus, "Eye fixations and recognition memory for pictures," *Cognitive Psychology*, vol. 3, no. 4, pp. 525–551, 1972.

[15] J. M. Henderson, "Human gaze control during real-world scene perception," *Trends in Cognitive Sciences*, vol. 7, no. 11, pp. 498–504, 2003.

[16] P. S. Holzman, L. R. Proctor, D. L. Levy, N. J. Yasillo, H. Y. Meltzer, and S. W. Hurt, "Eye-tracking dysfunctions in schizophrenic patients and their relatives," *Archives of General Psychiatry*, vol. 31, no. 2, pp. 143–151, 1974.

[17] D. Riby and P. J. Hancock, "Looking at movies and cartoons: eye-tracking evidence from williams syndrome and autism," *Journal of Intellectual Disability Research*, vol. 53, no. 2, pp. 169–181, 2009.

[18] A. Antervo, R. Hari, T. Katila, T. Ryhänen, and M. Seppänen, "Magnetic fields produced by eye blinking," *Electroencephalography and Clinical Neurophysiology*, vol. 61, no. 4, pp. 247–253, 1985.

[19] R. Hari, H. Salmellin, S. Tissari, M. Kajola, and V. Virsu, "Visual stability during eyeblinks." *Nature*, 1994.

[20] S. Y. Moon, J. J. Barton, S. Mikulski, F. E. Polli, M. S. Cain, M. Vangel, M. S. Hämäläinen, and D. S. Manoach, "Where left becomes right: a magnetoencephalographic study of sensorimotor transformation for antisaccades," *Neuroimage*, vol. 36, no. 4, pp. 1313–1323, 2007.

[21] A. A. Ioannides, P. B. Fenwick, and L. Liu, "Widely distributed magnetoencephalography spikes related to the planning and execution of human saccades," *The Journal of Neuroscience*, vol. 25, no. 35, pp. 7950–7967, 2005.

[22] J. E. McDowell, J. M. Kissler, P. Berg, K. A. Dyckman, Y. Gao, B. Rockstroh, and B. A. Clementz, "Electroencephalography/magnetoencephalography study of cortical activities preceding prosaccades and antisaccades," *Neuroreport*, vol. 16, no. 7, pp. 663–668, 2005.

[23] B. R. Cornwell, S. C. Mueller, R. Kaplan, C. Grillon, and M. Ernst, "Anxiety, a benefit and detriment to cognition: behavioral and magnetoencephalographic evidence from a mixed-saccade task," *Brain and Cognition*, vol. 78, no. 3, pp. 257–267, 2012.

[24] D. S. Manoach, A. K. Lee, M. S. Hämäläinen, K. A. Dyckman, J. S. Friedman, M. Vangel, D. C. Goff, and J. J. Barton, "Anomalous use of context during task preparation in schizophrenia: a magnetoencephalography study," *Biological Psychiatry*, vol. 73, no. 10, pp. 967–975, 2013.

[25] L. Hirvenkari, V. Jousmäki, S. Lamminmäki, V.-M. Saarinen, M. E. Sams, and R. Hari, "Gaze-direction-based meg averaging during audiovisual speech perception," *Frontiers in Human Neuroscience*, vol. 4, p. 17, 2010.

[26] R. W. Ditchburn and B. L. Ginsborg, "Vision with a stabilized retinal image," *Nature*, vol. 170, no. 4314, pp. 36–37, 1952.

[27] D. Robinson, "The mechanics of human smooth pursuit eye movement," *The Journal of Physiology*, vol. 180, no. 3, pp. 569–591, 1965.

[28] E. Matin, "Saccadic suppression: a review and an analysis." *Psychological Bulletin*, vol. 81, no. 12, p. 899, 1974.

[29] H. Barlow, "Eye movements during fixation," *The Journal of Physiology*, vol. 116, no. 3, pp. 290–306, 1952.

[30] J. Findlay, "A simple apparatus for recording microsaccades during visual fixation," *The Quarterly Journal of Experimental Psychology*, vol. 26, no. 1, pp. 167–170, 1974.

[31] D. A. Rosenbaum, *Human Motor Control*. Academic press, 2009.

[32] W. Becker and A. Fuchs, "Further properties of the human saccadic system: eye movements and correction saccades with and without visual fixation points," *Vision Research*, vol. 9, no. 10, pp. 1247–1258, 1969.

[33] J. Ross, M. C. Morrone, M. E. Goldberg, and D. C. Burr, "Changes in visual perception at the time of saccades," *Trends in Neurosciences*, vol. 24, no. 2, pp. 113–121, 2001.

[34] H. Deubel and W. X. Schneider, "Saccade target selection and object recognition: Evidence for a common attentional mechanism," *Vision Research*, vol. 36, no. 12, pp. 1827–1837, 1996.

[35] E. Kowler, E. Anderson, B. Dosher, and E. Blaser, "The role of attention in the programming of saccades," *Vision Research*, vol. 35, no. 13, pp. 1897–1916, 1995.

[36] J. E. Hoffman and B. Subramaniam, "The role of visual attention in saccadic eye movements," *Perception & Psychophysics*, vol. 57, no. 6, pp. 787–795, 1995.

[37] W. Becker and R. Jürgens, "An analysis of the saccadic system by means of double step stimuli," *Vision Research*, vol. 19, no. 9, pp. 967–983, 1979.

[38] A. Duchowski, *Eye Tracking Methodology: Theory and Practice*. Springer Science, 2003.

[39] SensoMotoric Instruments (SMI), *iView X^{TM} System Manual*.

[40] SR Research Ltd., *EyeLink 1000 User Manual*.

[41] L. R. Young and D. Sheena, "Survey of eye movement recording methods," *Behavior Research Methods & Instrumentation*, vol. 7, no. 5, pp. 397–429, 1975.

[42] A. Bulling, J. A. Ward, H. Gellersen, and G. Tröster, "Eye movement analysis for activity recognition," in *Proceedings of the 11th International Conference on Ubiquitous Computing*. ACM, 2009, pp. 41–50.

[43] M. Hämäläinen, R. Hari, R. J. Ilmoniemi, J. Knuutila, and O. V. Lounasmaa, "Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain," *Reviews of Modern Physics*, vol. 65, no. 2, p. 413, 1993.

[44] Y. C. Okada, A. Lahteenmäki, and C. Xu, "Experimental analysis of distortion of magnetoencephalography signals by the skull," *Clinical Neurophysiology*, vol. 110, no. 2, pp. 230–238, 1999.

[45] M. F. Bear, B. W. Connors, and M. A. Paradiso, *Neuroscience*. Lippincott Williams & Wilkins, 2007, vol. 2.

[46] R. Hari and R. Salmelin, "Magnetoencephalography: from squids to neuroscience: Neuroimage 20th anniversary special edition," *Neuroimage*, vol. 61, no. 2, pp. 386–396, 2012.

[47] P. Hansen, M. Kringelbach, and R. Salmelin, *MEG: An Introduction to Methods*. Oxford university press, 2010.

[48] Elekta Neuromag., *MaxFilter User's Guide*.

[49] J. M. Henderson, "Regarding scenes," *Current Directions in Psychological Science*, vol. 16, no. 4, pp. 219–222, 2007.

[50] G. R. Loftus and N. H. Mackworth, "Cognitive determinants of fixation location during picture viewing." *Journal of Experimental Psychology: Human Perception and Performance*, vol. 4, no. 4, p. 565, 1978.

[51] A. Friedman, "Framing pictures: the role of knowledge in automatized encoding and memory for gist." *Journal of Experimental Psychology: General*, vol. 108, no. 3, p. 316, 1979.

[52] J. M. Henderson, P. A. Weeks Jr, and A. Hollingworth, "The effects of semantic consistency on eye movements during complex scene viewing." *Journal of*

Experimental Psychology: Human Perception and Performance, vol. 25, no. 1, p. 210, 1999.

[53] M. S. Castelhano, M. L. Mack, and J. M. Henderson, "Viewing task influences eye movement control during active scene perception," *Journal of Vision*, vol. 9, no. 3, p. 6, 2009.

[54] T. Betz, T. C. Kietzmann, N. Wilming, and P. König, "Investigating task-dependent top-down effects on overt visual attention," *Journal of Vision*, vol. 10, no. 3, p. 15, 2010.

[55] C. A. Rothkopf, D. H. Ballard, and M. M. Hayhoe, "Task and context determine where you look," *Journal of Vision*, vol. 7, no. 14, p. 16, 2007.

[56] A. Oliva, A. Torralba, M. S. Castelhano, and J. M. Henderson, "Top-down control of visual attention in object detection," in *Image processing, 2003. icip 2003. proceedings. 2003 international conference on*, vol. 1. IEEE, 2003, pp. I–253.

[57] D. J. Parkhurst and E. Niebur, "Scene content selected by active vision," *Spatial Vision*, vol. 16, no. 2, pp. 125–154, 2003.

[58] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, no. 11, pp. 1254–1259, 1998.

[59] L. Itti and C. Koch, "A saliency-based search mechanism for overt and covert shifts of visual attention," *Vision Research*, vol. 40, no. 10, pp. 1489–1506, 2000.

[60] D. Parkhurst, K. Law, and E. Niebur, "Modeling the role of salience in the allocation of overt visual attention," *Vision Research*, vol. 42, no. 1, pp. 107–123, 2002.

[61] E. Birmingham, W. F. Bischof, and A. Kingstone, "Saliency does not account for fixations to eyes within social scenes," *Vision Research*, vol. 49, no. 24, pp. 2992–3000, 2009.

[62] M. Cerf, E. P. Frady, and C. Koch, "Faces and text attract gaze independent of the task: Experimental data and computer model," *Journal of Vision*, vol. 9, no. 12, p. 10, 2009.

[63] M. Bindemann, A. M. Burton, I. T. Hooge, R. Jenkins, and E. H. De Haan, "Faces retain attention," *Psychonomic Bulletin & Review*, vol. 12, no. 6, pp. 1048–1053, 2005.

[64] M. Sams, J. Hietanen, R. Hari, R. Ilmoniemi, and O. V. Lounasmaa, "Face-specific responses from the human inferior occipito-temporal cortex," *Neuroscience*, vol. 77, no. 1, pp. 49–55, 1997.

[65] E. Halgren, T. Raij, K. Marinkovic, V. Jousmäki, and R. Hari, "Cognitive response profile of the human fusiform face area as determined by meg," *Cerebral Cortex*, vol. 10, no. 1, pp. 69–81, 2000.

[66] J. Liu, A. Harris, and N. Kanwisher, "Stages of processing in face perception: an meg study," *Nature Neuroscience*, vol. 5, no. 9, pp. 910–916, 2002.

[67] T. Tanskanen, R. Näsänen, T. Montez, J. Päällysaho, and R. Hari, "Face recognition and cortical responses show similar sensitivity to noise spatial frequency," *Cerebral Cortex*, vol. 15, no. 5, pp. 526–534, 2005.

[68] B. Rossion and S. Caharel, "Erp evidence for the speed of face categorization in the human brain: Disentangling the contribution of low-level visual cues from face perception," *Vision Research*, vol. 51, no. 12, pp. 1297–1311, 2011.

[69] D. Purves, *Neuroscience*, 5th ed. Sunderland, Mass: Sinauer Associates, 2012.

[70] G. McCarthy, A. Puce, J. C. Gore, and T. Allison, "Face-specific processing in the human fusiform gyrus," *Journal of Cognitive Neuroscience*, vol. 9, no. 5, pp. 605–610, 1997.

[71] V. Willenbockel, J. Sadr, D. Fiset, G. O. Horne, F. Gosselin, and J. W. Tanaka, "Controlling low-level image properties: the shine toolbox," *Behavior Research Methods*, vol. 42, no. 3, pp. 671–684, 2010.

[72] J. W. Peirce, "Psychopy—psychophysics software in python," *Journal of Neuroscience Methods*, vol. 162, no. 1, pp. 8–13, 2007.

[73] F. Cristino, S. Mathôt, J. Theeuwes, and I. D. Gilchrist, "Scanmatch: A novel method for comparing fixation sequences," *Behavior Research Methods*, vol. 42, no. 3, pp. 692–700, 2010.

[74] S. W. Kuffler, "Discharge patterns and functional organization of mammalian retina," *Journal of Neurophysiology*, vol. 16, no. 1, pp. 37–68, 1953.

[75] D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *The Journal of Physiology*, vol. 160, no. 1, p. 106, 1962.

[76] P. J. Burt and E. H. Adelson, "The laplacian pyramid as a compact image code," *Communications, IEEE Transactions on*, vol. 31, no. 4, pp. 532–540, 1983.

[77] M. Jiang, J. Xu, and Q. Zhao, "Saliency in crowd," in *Computer Vision–ECCV 2014*. Springer, 2014, pp. 17–32.

[78] N. Kriegeskorte, M. Mur, and P. Bandettini, "Representational similarity analysis–connecting the branches of systems neuroscience," *Frontiers in Systems Neuroscience*, vol. 2, 2008.

[79] J. V. Haxby, M. I. Gobbini, M. L. Furey, A. Ishai, J. L. Schouten, and P. Pietrini, "Distributed and overlapping representations of faces and objects in ventral temporal cortex," *Science*, vol. 293, no. 5539, pp. 2425–2430, 2001.

[80] R. M. Cichy, D. Pantazis, and A. Oliva, "Resolving human object recognition in space and time," *Nature Neuroscience*, vol. 17, no. 3, pp. 455–462, 2014.

[81] K. Guo, S. Mahmoodi, R. G. Robertson, and M. P. Young, "Longer fixation duration while viewing face images," *Experimental Brain Research*, vol. 171, no. 1, pp. 91–98, 2006.

[82] J. J. Barton, N. Radcliffe, M. V. Cherkasova, J. Edelman, and J. M. Intriligator, "Information processing during face recognition: The effects of familiarity, inversion, and morphing on scanning fixations," *Perception*, vol. 35, pp. 1089–1105, 2006.

A Locations of faces in the images

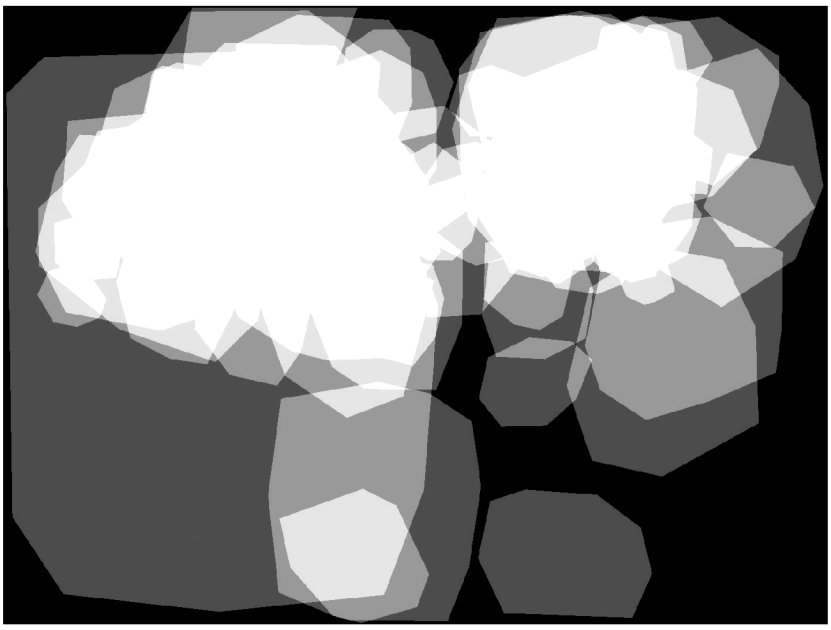


Figure A1: Locations of faces in images containing 1 face.

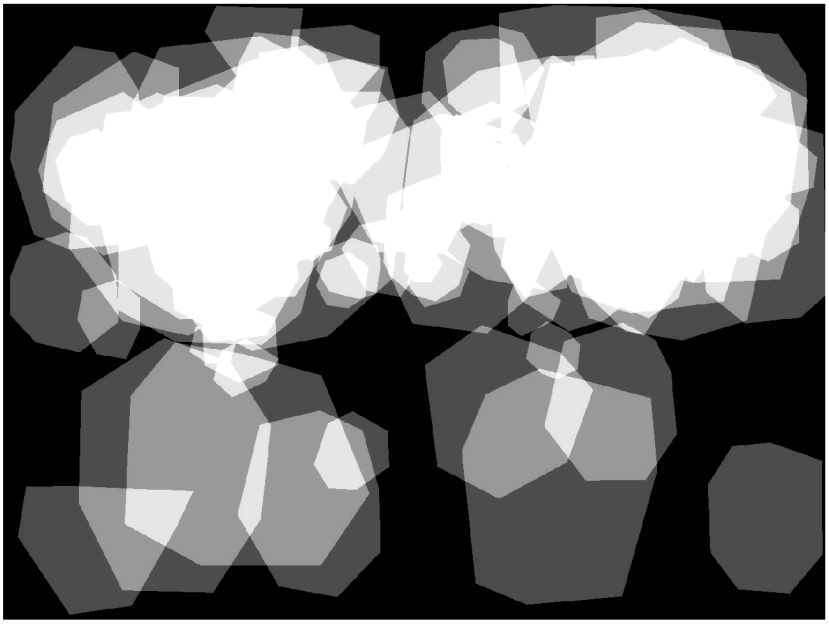


Figure A2: Locations of faces in images containing 2 faces.

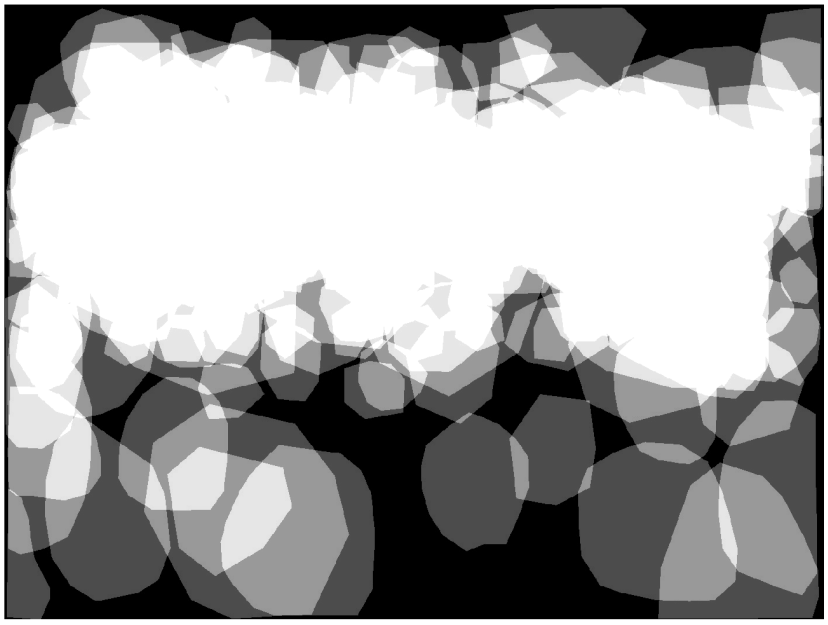


Figure A3: Locations of faces in images containing 3 or more faces.

B The PsychoPy script

Listing 1: Experiment setup script

```
# -*- coding: utf-8 -*-
from psychopy import visual, logging, event, core
from psychopy.iohub import (EventConstants, ioHubExperimentRuntime, module_directory,
                            getCurrentDateTimeString)
import os, time, parallel, numpy, codecs, random, serial




def sendTrigger(data):

    #output triggers
    port.setData(data)

    # make sure the trigger pulses are at least 10 ms in duration; otherwise they might
     be missed.
    time.sleep(0.01)

    # Reset the trigger lines
    port.setData(0)



def sendTrackerMessage(msg):

    tracker.sendMessage(msg, time_offset=None)



class ExperimentRuntime(ioHubExperimentRuntime):
    """
    Create an experiment using psychopy and the ioHub framework by extending
    the ioHubExperimentRuntime class and implementing the run() method.
    """


    def run(self,*args):


        ### Log file

        # Save log to a file
        exp_time = time.strftime("%d%m%Y") + '_' + time.strftime('%H%M')
        logDirectory = os.getcwd().replace("\\", "/")  + '/logDat_' + exp_time + '.log'
```

```
44          logDat=logging.LogFile(logDirectory, level=logging.INFO, filemode='a')
45
46
47          ### Fetch images and questions
48
49          # Read questions from a text file and save them to an array
50          questions = []
51          quest_txt = codecs.open('questions.txt', 'r', encoding='utf-8')
52          for line in quest_txt:
53              questions.append(unicode(line.strip()))
54          quest_txt.close()
55
56          # Image directory
57          imgdir='D:/Users/Kaisu/ImagesLumMatch/'
58
59          files=[]
60          for file in os.listdir(imgdir):
61              # Find all .jpg or .png images
62              if file.lower().endswith(".jpg")  or file.lower().endswith(".png") :
63                  # Load file to the list of files
64                  files.append(file)
65
66
67
68          ### Create image and question order
69
70          # Randomize the file list
71          random.shuffle(files)
72
73          # How many images used in one part
74          p = 50
75
76          # Loop the questions to get as many questions as there are images in one part
77          questions_finale=[]
78          qi = 0
79          q = qi
80          while q < p+25:
81              questions_finale.append(questions[qi])
82              if (qi+1) == len(questions):
83                  qi=0
84              else:
85                  qi+=1
86              q+=1
87
88          # In how many parts the experiment is divided
89          parts = 4
90
91
92
```

```
93          ### Devices, initialization
94
95          # Short-cuts to the devices
96          global tracker
97          tracker=self.hub.devices.tracker
98          display=self.hub.devices.display
99          kb=self.hub.devices.kb
100         mouse=self.hub.devices.mouse
101
102         KEYBOARD_PRESS=EventConstants.KEYBOARD_PRESS
103
104         # Hide the 'system mouse cursor' during the experiment.
105         mouse.setSystemCursorVisibility(False)
106
107         # Eye tracker calibration etc
108         tracker.runSetupProcedure()
109
110
111
112         ### Setup the Window etc.
113
114         # Create window
115         win=visual.Window(size=(1400, 1050), pos=None, units='pix', fullscr=True,
    screen=1, monitor="default", allowGUI=False)
116
117         # Create fixation cross
118         crossImg=visual.ImageStim(win, image=os.getcwd()+'\cross.png', units='pix', pos
    =(0.0, 0.0), size=(50,50), flipHoriz=False, flipVert=False, name=file, autoLog=
    False)
119
120         squareImg=visual.ImageStim(win, image=os.getcwd()+'\square.png', units='pix',
    pos=(-700.0, -525.0), size=(50,50), flipHoriz=False, flipVert=False, name=file,
    autoLog=False)
121
122         testquest = visual.TextStim(win, autoLog=False, units='pix', height=40)
123
124         timer = core.Clock()
125         timer2 = core.Clock()
126
127
128         for a in range(parts):
129
130             ### Load images for one part
131
132             # Make sure the images have time to be loaded
133             timing = core.StaticPeriod(screenHz=60)
134             # Start a period of 20.0s
135             timing.start(20.0)
136
```

```
137             # Empty list for images
138             images = []
139 
140             # Check that index does not go out of range (less images in the last part)
141             if((len(files)-(a*p)) < p):
142                 r = len(files)-(a*p)
143             else:
144                 r=p
145 
146             # Load images used in one part
147             for f in range(r):
148                 file=files[(f+(a*p))]
149 
150                 # Flip 25 images
151                 if f < 25:
152                     images.append(visual.ImageStim(win, image=imgdir+file , units='pix', pos=(0.0, 0.0), size=(1400, 1050), flipHoriz=False, flipVert=True, name='FLIPPED_' + file , autoLog=False))
153                     images.append(visual.ImageStim(win, image=imgdir+file , units='pix', pos=(0.0, 0.0), size=(1400, 1050), flipHoriz=False, flipVert=False, name=file , autoLog=False))
154                 else:
155                     images.append(visual.ImageStim(win, image=imgdir+file , units='pix', pos=(0.0, 0.0), size=(1400, 1050), flipHoriz=False, flipVert=False, name=file , autoLog=False))
156 
157             # Randomize questions
158             random.shuffle(questions_finale)
159 
160             # Shuffle images (also flipped images are randomized)
161             random.shuffle(images)
162 
163             # Finish 20.0s
164             timing.complete()
165 
166             #
167             testquest.setText('Kerro kun olet valmis aloittamaan', None)
168             testquest.draw()
169             win.flip()
170 
171             # Run calibration etc again
172             keys=event.waitKeys()
173             while 'e' in keys:
174                 tracker.runSetupProcedure()
175 
176                 testquest.draw()
177                 win.flip()
178                 keys=event.waitKeys()
179 
```

```
180
181             # Start Recording Eye Data
182             tracker.setRecordingState(True)
183
184             # Image 1,0 s = 60 frames
185             # Question 4,0 s = 240 frames
186             i_dur = 60
187             q_dur = 240
188             dur=i_dur+q_dur
189
190             for image, question in zip(images, questions_finale):
191
192
193                 c_duration = random.randint(130,210)   # Randomize the duration of
    fixation cross
194
195                 yes=False
196                 no=False
197                 answered=False
198
199                 crossImg.draw()
200                 squareImg.draw()
201                 if images.index(image) == 0:
202                     win.logOnFlip(msg= 'Part %i started\n' %(a+1), level=logging.INFO)
203                     win.callOnFlip(sendTrackerMessage, 'Part %i started\n' %(a+1))
204                     win.callOnFlip(sendTrigger, (a+2))
205                 win.flip()
206                 timing.start(c_duration/100.0)
207                 image.draw()
208                 timing.complete()
209
210
211                 for frameN in range(dur+1):
212
213                     if event.getKeys(['escape']):
214                         win.close()
215                         tracker.setRecordingState(False)
216                         tracker.setConnectionState(False)
217                         core.quit()
218
219
220                     # Draw image (and cross) and save the point of time of image
    stimulus
221                     if frameN == 0:
222                         crossImg.draw()
223                         win.logOnFlip(msg= 'IMAGE: ' + image.name , level=logging.INFO)
224                         win.callOnFlip(sendTrigger, 1)
225                         win.callOnFlip(sendTrackerMessage, image.name)
226                         timer.reset()
```

```
227
228                     elif frameN < i_dur:
229                         image.draw()
230                         crossImg.draw()
231
232                     elif frameN == i_dur:
233                         testquest.setText(question, None)
234                         testquest.draw()
235                         win.logOnFlip(msg= 'QUESTION: ' + question.encode('ascii',
    errors='replace'), level=logging.INFO)
236                         win.callOnFlip(sendTrackerMessage, 'Question: ' + question.
    encode('ascii', errors='replace'))
237                         timer2.reset()
238                         ser.flushInput()
239
240                     elif frameN == (i_dur+1):
241                         itime = timer.getTime()
242                         testquest.draw()
243                         win.logOnFlip(msg= 'Image shown: ' + str(itime), level=logging.
    INFO)
244
245                     elif frameN < dur:
246                         if answered:
247                             if yes:
248                                 testquest.setText(question + (u'\n Vastasit kylla'),
    None)
249                             elif  no:
250                                 testquest.setText(question + (u'\n Vastasit ei'), None)
251                         testquest.draw()
252                         x=ser.read(1)
253                         if (str(x.encode('hex')) == '01') and not answered:
254                             answered=True
255                             yes=True
256                         if (str(x.encode('hex')) == '02') and not answered:
257                             answered=True
258                             no=True
259
260
261                     else:
262                         qtime = timer2.getTime()
263                         win.logOnFlip(msg= 'Question shown: ' + str(qtime), level=
    logging.INFO)
264                         if yes:
265                             win.logOnFlip(msg= 'Answer: YES', level=logging.INFO)
266                         elif no:
267                             win.logOnFlip(msg= 'Answer: NO', level=logging.INFO)
268                         else:
269                             win.logOnFlip(msg= 'Answer: SKIP', level=logging.INFO)
270
```

```
                    flip_time = win.flip()


            # Stop recording eye data.
            #
            tracker.setRecordingState(False)


        win.close()


        # Disconnect the eye tracking device.
        #
        tracker.setConnectionState(False)

        ser.close()


# Meg triggers
port = parallel.Parallel(port=0)
port.setData(0)

# Serial port for responses
ser = serial.Serial(0, 115200, timeout=0)  # open first serial port


######## Launch the Experiment #######

runtime=ExperimentRuntime(module_directory(ExperimentRuntime.run), "experiment_config.
    yaml")
runtime.start()
```