Processing fast event-related fMRI data for Artificial Neural Network Applications

by

Hannah Terborg

Submitted to attain the academic degree
Bachelor of Science
in Cognitive Science

University of Osnabrück
Faculty of Human Sciences

First supervisor: Dr. Teresa Cheung
Secondary supervisor: Pascal Nieters, M. Sc.

Submission date: 31.03.2021

Matriculation number: 969021
Semester: 9

Abstract

Recently artificial neural networks (ANNs) have been used to perform classification tasks on data from functional magnetic resonance imaging (fMRI). ANNs often require data-label pairs as input. These can easily be created from fMRI data with a block design. However, labelling fMRI scans is challenging when using fast event-related fMRI data due to the specific characteristics of the blood oxygen level dependent (BOLD) signal, such as its 6-9 seconds delay. The herein persented study proposes a processing pipeline to create data-label pairs from fast event-related fMRI data by labelling each scan and subsequently optimizing this labelling in a brute force manner. This method incorporates the BOLD signal delay into the scan labelling and excludes scans that cannot be given a definite label. The study uses data from a face paradigm by Henson et al. (2003). The trial labelling resulting from the application of the processing pipeline is tested by training a convolutional neural network proposed by Vu et al. (2020). As a results, this study successfully creates data-label pairs from fast event-related fMRI data. The CNN trained on this data further uses face-selective regions in the classification process, as saliency maps reveal. Nevertheless, the study reveals a lack of ground truth when working with testing data from fast event-related fMRI paradigms. Overall, the presented study evidences that using fast event-related fMRI data-label pairs for artificial neural network training is possible.

Acknowledgements

First and foremost, I have to thank my supervisors Dr. Teresa Cheung and Pascal Nieters for their assistance and dedicated involvement in this project.

Further, I like to thank WestGrid (www.westgrid.ca) and Compute Canada Calcul Canada (www.computecanada.ca) for their computational resources.

I would also like to thank everyone who provided me with their support throughout this project by proofreading, giving advice or encouragements or sharing their experience.

Contents

List of Acronyms

ANN	artificial neural network
ASD	autism spectrum disorder
BOLD	blood oxygen level dependent
CBF	cerebral blood flow
CBV	cerebral blood volume
CNN	convolutional neural network
CPU	central processing unit
DICOM	Digital Imaging and Communications in Medicine
EEG	electroencephalography
FFA	fusiform face area
fMRI	functional magnetic resonance imaging
GLM	general linear model
GPU	graphic processing unit
HRF	hemodynamic response function
ISI	inter-stimulus interval
ITI	inter-trial interval
MEG	magnetoencephalography
MNI	Montreal Neurological Institute
MNIST	modified National Institute of Standards and Technology
MPRAGE	magnetization prepared rapid gradient echo
MRI	magnetic resonance imaging
NIfTI	Neuroimaging Informatics Technology Initiative
NOS	non optimal shift
NSMD	negative squared mean difference
OFA	occipital face area
OS	optimal shift
PET	positron emission tomography
PSC	percent signal change
RAM	random-access memory
ROI	region of interest
sMRI	structural magnetic resonance imaging
SPM	statistical parameter mapping
STS	superior temporal sulcus
TE	echo time
TR	repetition time

List of Figures

List of Tables

1 Introduction

What is happening inside of your brain just in this moment? What is the person next to you seeing, maybe the face of a loved one? Using a combination of machine learning and neuroimaging methods might be able to answer these questions and make "mind reading" possible.

Artificial neural networks are a machine learning technique that has gained popularity over the past years. Recently, they have been applied to neuroimaging data to make inferences about brain states. This approach could be used to aid diagnosis of neurological and psychological conditions in the future. An artificial neural network could, thereby, possibly classify a neurological disease earlier than possible with the current assessments methods. It might further reveal the essential neurological difference between a population with a certain condition and their neurotypical counterpart.

One popular neuroimaging method is functional magnetic resonance imaging (fMRI) which produces 3-dimensional images of the brain. Studies which use fMRI data and fMRI-based brain decoding tasks, such as classification, usually use an experimental design called block design (Lee et al., 2017). In this design multiple experimental stimuli of one condition are presented together in a block. Block design, however, has limitations as it cannot be used to investigate all experimental paradigms. Therefore, many studies have used the more flexible fast event-related design instead (Henson et al., 2003). In a fast event-related fMRI study stimuli are not presented as blocks of separate conditions but as mixed individual events.

Artificial neural networks learn from example. For that they need data-label pairs consisting of the data, e.g an image, and its label, i.e what the image shows. An example for such a network is the convolutional neural network by Vu et al. (2020). The data-label pairs in this case consist of a brain scan and a label, which corresponds to the condition of a seen stimuli. These data-label pairs can easily be created from fMRI data with a block experimental design. In this case, the label is the condition of the block of stimuli a participant has seen and the data one fMRI scan acquired in the block, respectively. The properties of fast event-related designs, such as the mix of stimuli and, thereby, conditions presented in a short amount of time, however, make creation of data-label pairs much more complicated.

Further, as functional MRI acquisition is challenging and comes with high costs and low sample sizes (Yin et al., 2020), as much data as possible should be repurposed, when dealing with data demanding machine learning applications, like artificial neural networks This includes data of already existing studies, as well as data from future studies. In line with this, the herein presented thesis is concerned with creating adequate data-label pairs from a fast event-related fMRI design. More

specifically the fast event-related fMRI data from a face paradigm published by Henson et al. (2003) will be used,

Henson et al. (2003) have published a multi-modal data set for a face paradigm, which includes fast event-related fMRI data. Based on their data the study presented in this thesis will attempt to create data-label pairs which serve as input to the artificial neural network presented by Vu et al. (2020) to attempt classification of fMRI volumes into the categories "face" and "noface", corresponding to the stimuli seen by the participant. For this attempt a processing pipeline is proposed. The study will pursue to answer two main research questions: Firstly, whether or not it is possible to create data-label pairs from fast event-related fMRI data for artificial neural network applications and, secondly, how a convolutional neural network performs with such an input.

This thesis will at first explain the neuroimaging method fMRI and the blood oxygen level dependent (BOLD) signal. Further, block and event-related experimental designs will be introduced in detail and their differences presented. Afterwards, face processing in the human brain will be covered. Lastly, artificial neural networks and their applications in neuroimaging will be demonstrated. This part will also include the introduction to the convolutional neural network (CNN) by Vu et al. (2020).

The following chapter will more closely establish the study conducted by specifying the two main research questions. Then the methodology section follows, which will present the data set from Henson et al. (2003), the computational method used, as well as the sub-steps of the processing pipeline. These sub-steps correspond to different algorithms used and include preprocessing in SPM, a shifting and labelling algorithm, an optimization algorithm,a normalization step, the creation of data sets and, lastly, the application of the artificial neural network.

The results of this pipeline and the individual sub-steps will be demonstrated in the fourth chapter. These results will be debated in the discussion, where also limitations of the study will be reviewed and recommendations for future studies expressed.

2 Theoretical Background

2.1 Functional magnetic resonance imaging (fMRI)

Functional magnetic resonance imaging (fMRI) is a widely used non-invasive imaging method for studying the brain (Amaro & Barker, 2006; Glover, 2011; Kim & Bandettini, 2012; Logothetis & Wandell, 2004), in which a magnetic field is used to create brain images (Huettel et al., 2008, Chapter 1). Functional MRI is based on magnetic resonance imaging (MRI) (Glover, 2011), also referred to as structural magnetic resonance imaging (sMRI), which is used to investigate anatomical features of the brain (cf. Huettel et al., 2008, Chapter 4), whereas functional MRI indirectly measures neural activity (Logothetis & Wandell, 2004) over time (Glover, 2011). Following the first measurements on humans taken in 1992, fMRI has been mostly utilized in research but has also found application in clinical environments, for example in the preparation of brain surgeries (Glover, 2011).

2.1.1 The BOLD signal

The blood oxygen level-dependent (BOLD) signal was first discovered by Ogawa et al. and reflects blood oxygen level-dependent changes in the brain (Logothetis & Wandell, 2004; Ogawa et al., 1990). The BOLD signal is based on the specific magnetic properties of hemoglobin (Amaro & Barker, 2006; Glover, 2011; Huettel et al., 2008; Logothetis & Wandell, 2004). Oxygenated hemoglobin (oxyhemoglobin) is diamagnetic. However, when the oxyhemoglobin releases its oxygen molecule it becomes deoxygenated hemoglobin (deoxyhemoglobin), which is paramagnetic. Functional MRI is sensitive to these property differences and picks up changes in the ratio of oxyhemoglobin to deoxyhemoglobin as a change in BOLD signal intensity. As a consequence, regions with high amounts of deoxyhemoglobin appear in the resulting images as darker, those with high amounts of oxyhemoglobin as lighter (Ogawa et al., 1990). In this way, deoxyhemoglobin acts as a natural, intrinsic contrast agent (Glover, 2011; Ogawa et al., 1990).

The amount of deoxyhemoglobin at a specific location is driven by hemodynamics, mainly the cerebral blood flow (CBF) and cerebral blood volume (CBV), as well as oxygen consumption, which is a metabolic aspect (Amaro & Barker, 2006; Logothetis & Wandell, 2004). Neural activity influences these hemodynamic and metabolic factors. This mechanism is termed neurovascular coupling (Hillman, 2014): Neural activity requires energy in the form of adenosine triphosphate (ATP), leading to an increased oxygen consumption, this in turn influences the demand for oxygen and thereby increases CBF (Glover, 2011; Logothetis & Wandell, 2004). An increase in CBF leads to an excess amount of oxygen causing an increase in the BOLD signal.

The exact reasons behind this oversupply are, however, not yet fully understood (Hillman, 2014; Logothetis & Wandell, 2004).

What is commonly referred to as fMRI specifically examines the BOLD signal, its more precise name is thus BOLD fMRI (Glover, 2011).

The hemodynamic response function

To model or interpret the BOLD signal, a characterizing hemodynamic response function (HRF) can be utilized (Hillman, 2014). Logothetis and Wandell (2004) define the HRF as follows:

> The time course of the human BOLD response to a brief stimulus, the temporal impulse response function, is often called the hemodynamic response function (HRF). (Logothetis & Wandell, 2004)

This means that the HRF depicts changes in the measured signal over time following neural activity (Huettel et al., 2008, Chapter 7). Figure 2.1 shows SPM's representation of a HRF upon hypothetical presentation of a brief stimulus. The general form of the HRF can be described as follows: The BOLD response begins roughly two seconds after the stimulus was presented, rising to a peak or, in the case of multiple stimuli, to a plateau, before returning to its baseline (Logothetis & Wandell, 2004). The time to peak of the HRF is specified differently between sources ranging from as early as 3 – 5 seconds (Amaro & Barker, 2006; Hillman, 2014), over 5 – 6 seconds (Glover, 2011), up to 6 – 9 seconds (Logothetis & Wandell, 2004) after stimulus onset. It further needs to be noted that the HRF is not always the same, but that it differs between individuals (Logothetis & Wandell, 2004), among different areas across the cortex (Drew, 2019; Logothetis & Wandell, 2004), depending on the task (Logothetis & Wandell, 2004) and the hormonal state (Drew, 2019). Moreover, there are proposals that the HRF possesses an initial dip after stimulus presentation or an undershoot after peaking (Hillman, 2014; Huettel et al., 2008; Logothetis & Wandell, 2004).

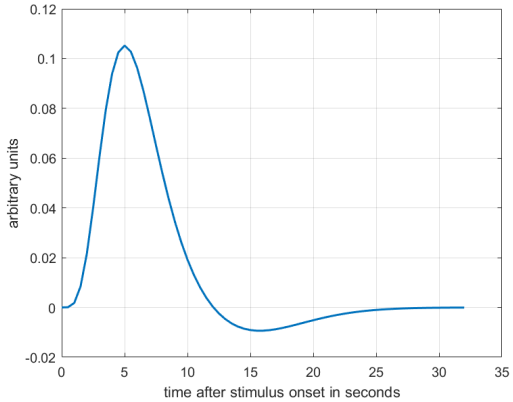


Figure 2.1: SPMs canonical HRF.

2.1.2 One fMRI scan

Functional MRI creates volumetric images, which means a single fMRI image/scan is 3-dimensional. Comparable to the notion of pixels in a two dimensional image a fMRI volume consists of volumetric pixels, so-called voxels (Huettel et al., 2008, Chapter 1). To investigate a specific part of a brain volume-specific sections can be looked at: axial, sagittal and coronal slices. This is comparable to looking at the brain from above, the side and the front, respectively (Huettel et al., 2008, Glossary).

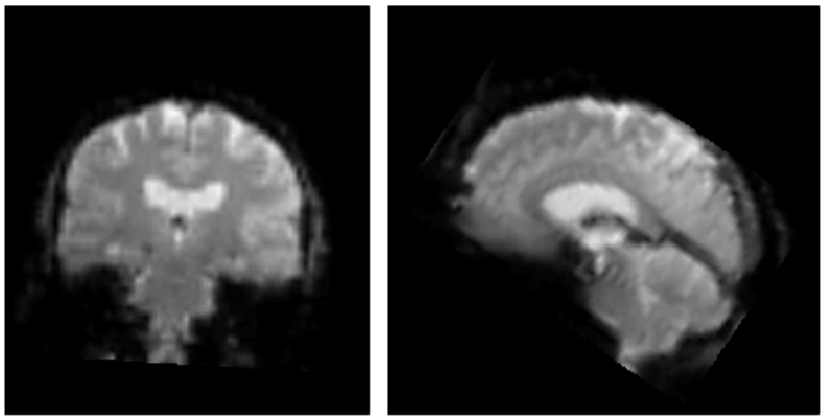


Figure 2.2: Example of a raw brain scan from Henson et al. (2003), coronal and sagittal slice at origin.

In Figure 2.2, a coronal and a sagittal slice from a single raw fMRI scan are shown. Due to the different magnetic properties of individual types of tissues the general structure of the brain's anatomy, such as white and grey matter, can be discerned. Despite this participant's engagement in a visual task during the scan, pronounced visual cortex activity cannot be made out by the naked eye. This is due to the relatively small range of BOLD signal changes, amounting to 1 – 10 % depending on the scanner used (Amaro & Barker, 2006). Another part of the image is noise, which is signal that is unrelated to neural activity. This noise can have various causes, such as physiological processes like breathing and heartbeat (Drew, 2019) or can have origins related to equipment (Glover, 2011). The relationship between noise and signal is called signal-to-noise ratio (Logothetis & Wandell, 2004) and is dependent on imaging parameters such as the magnetic field strength of the machine (Glover, 2011).

2.1.3 Critique

Over time, critique has been expressed in regard to BOLD functional magnetic resonance imaging (fMRI). Drew (2019) indicates that the linkage between neural activity and hemodynamic response is not as strong in all areas as in the primary sensory cortex, where it first had been researched, and advises caution in inferring neural activity from hemodynamic responses. Glover (2011) points out the low tem-

poral resolution as well as problems in studying ventral, temporal prefrontal cortex regions due to differences in magnetic susceptibility. After measuring impossible brain activity in a dead salmon, Bennett et al. (2009) impudently demonstrated the risk of false positives in fMRI and advised to be cautious when doing fMRI analyses and interpreting related statistics. This excerpt of still remaining problems as well as the existence of still unanswered questions about different aspects of fMRI imaging (Amaro & Barker, 2006; Hillman, 2014; Logothetis & Wandell, 2004), for example in relation to the details of neurovascular coupling (Hillman, 2014), show a great number of research possibilities in this field. Moreover, fMRI has proven itself as a powerful tool for making inferences about brain activity (Amaro & Barker, 2006; Kim & Bandettini, 2012; Logothetis & Wandell, 2004) as it has an overall high spatial resolution (Logothetis & Wandell, 2004) but a limited temporal resolution (Glover, 2011). Compared to positron emission tomography (PET), which uses a radioactive tracer, fMRI is not only less invasive, less expensive but also has a higher temporal and spatial resolution. It is, therefore, often the method of choice to locate where something in the brain is happening. The method of choice to investigate when something is happening in the brain are electroencephalography (EEG) and magnetoencephalography (MEG) due to their superior temporal resolution.

2.2 Experimental designs

Scientists designing a fMRI experiment have to decide how they want to present the stimuli of different conditions to the participants. Two major experimental designs can be distinguished: block design and event-related design.

2.2.1 Block design

In a block design, stimuli are presented in sequential blocks alternating between conditions. This means that after each block of stimuli of the experimental condition, e.g. a motor task, an equally long block of a control task, e.g. resting, follows (Amaro & Barker, 2006; Huettel, 2012; Rosen et al., 1998). Figure 2.3 shows an example of a block design with three conditions with a difference in the strength of the hemodynamic response. In Figure 2.3 a) the blocking of the stimuli can be seen, in b) the BOLD response upon the individual trials is drawn while c) shows the combination of the individual responses to an overall BOLD signal.

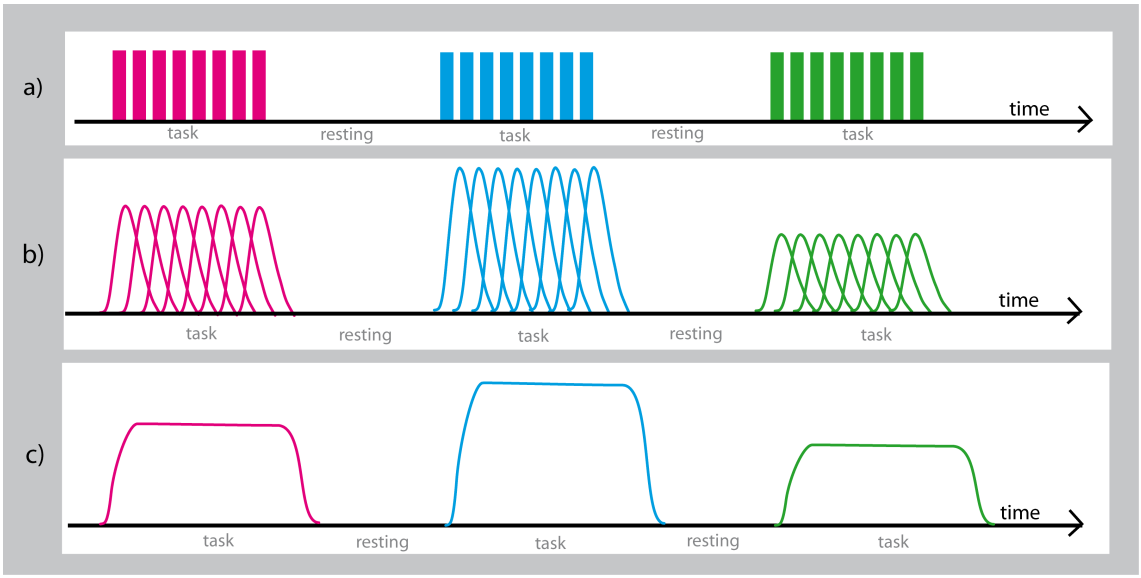


Figure 2.3: Schematic drawing of a block design with 3 conditions with differences in strength of the BOLD signal, a) blocking of trials b) BOLD response to every trial c) combined BOLD response.

The block design was dominantly adopted in the very first fMRI experiments (Amaro & Barker, 2006; Bennett et al., 2009; Huettel, 2012). This is due to its usage with the older PET. In PET a radioactive tracer is used to detect blood flow changes lasting over one minute (Amaro & Barker, 2006) making a blocked design necessary.

An advantage of block design, besides its simplicity (Huettel, 2012), is the size of the signal. Due to the blocking of stimuli, the associated hemodynamic changes are maximized (Amaro & Barker, 2006; Huettel, 2012). However, it has major neuropsychological drawbacks (Amaro & Barker, 2006), for example the repetition of the same stimulus type can lead to priming or habituation effects (Lee et al., 2017). The participant can therefore already anticipate the next stimulus. This leads to the experimental design and, therefore, the researched paradigms to be limited in block design fMRI experiments (Huettel et al., 2008, Chapter 9).

2.2.2 Event-related design

Event-related designs became popular in the mid-90s (Amaro & Barker, 2006) after the first event-related design was conducted in 1992 (Huettel, 2012). In an event-related design individual trials are no longer blocked but appear as single events often in a randomized order (Huettel et al., 2008, Chapter 9). Within this design, one can distinguish between slow and fast event-related design.

Slow event-related design

In a slow event-related design individual trials, so-called events, are presented around 12 – 16 seconds apart from each other (Burock et al., 1998; Huettel, 2012). This gap

between stimuli is called inter-stimulus interval (ISI) (Amaro & Barker, 2006). It is implemented to minimize the overlap of the hemodynamic responses to the individual stimuli (Huettel, 2012). If the ISI or alternatively the inter-trial interval (ITI) differs among trials, this is referred to as jittering (Huettel et al., 2008, Glossary). In this way different time points are sampled, leading to an improvement in temporal and spatial resolution (Amaro & Barker, 2006).

This design allows for exploration of temporal characteristics of the BOLD response (Amaro & Barker, 2006) while also having some of the same limitations of block design as the choice of experimental paradigms is limited (Burock et al., 1998). Slow event-related designs further suffer from the smaller number of trials presented over a certain time span (Burock et al., 1998).

Fast event-related design

In a fast or rapid event-related design the ISI is shorter than the hemodynamic response to the previous stimuli (Amaro & Barker, 2006), which results in overlap of the individual event's hemodynamic responses (Huettel, 2012). Figure 2.4 shows a schematic drawing of a fast event-related design with a constant ISI, where a) describes the presentation order and condition of the stimulus. In Figure 2.4 b) the BOLD response to the individual stimuli is indicated while c) shows the linear addition of these. Figure 2.4 thereby shows the overlap of BOLD responses to the individual trials which sum up to a complex BOLD signal.

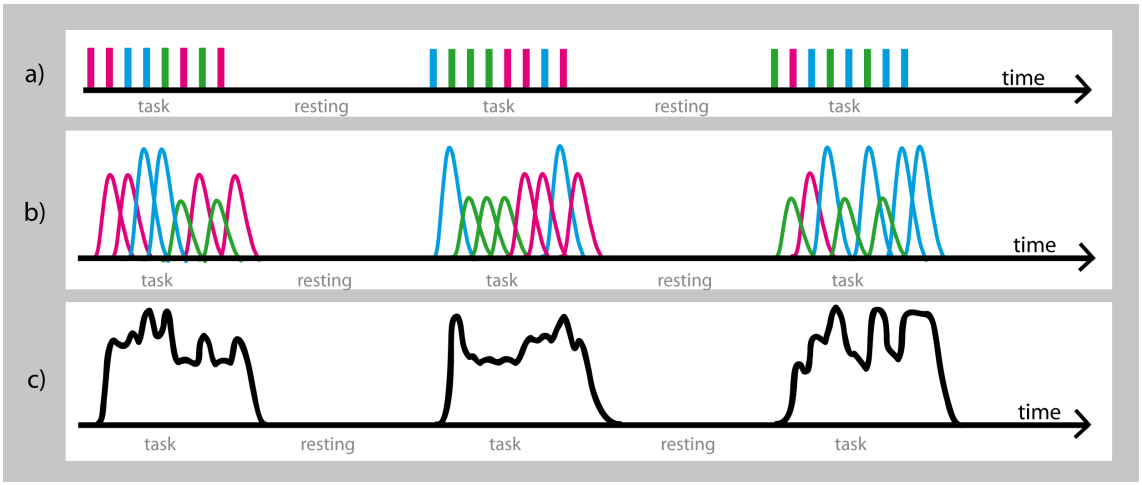


Figure 2.4: Schematic drawing of a fast event-related design with three conditions, with differences in the strength of the BOLD signal: a) indicates the presentation of the stimuli in individual trials, b) shows the BOLD signal for each stimuli, c) is the overall estimated BOLD response.

This overlap of hemodynamic responses influences the overall BOLD signal. Therefore, the linearity of the hemodynamic response was investigated questioning whether or not two hemodynamic responses add up linearly to their sum. One problem in

answering this question lies in the lack of clarity regarding whether or not the neural response itself is linearly additive over time (Rosen et al., 1998).

> The most parsimonious explanation for all of these findings is that the basic transformation between the summation of neuronal events and the BOLD response is approximately linear, at least with presentation rates typically used in fMRI studies. (Rosen et al., 1998)

This implies that the responses to multiple stimuli combine in a roughly linear manner. However, some non-linearities have been reported. Nonetheless, these seem to be small enough to justify the usage of event-related design in general but have influenced its analysis methods (Huettel, 2012).

The fast event-related design provides great flexibility (Amaro & Barker, 2006; Huettel, 2012; Rosen et al., 1998), allowing the experimental designs to match those of behavioural and electrophisological studies (Burock et al., 1998), as well as accessing practice effects (Amaro & Barker, 2006). Besides this, it is possible to evaluate and sort individual trials based on the participant's response (Amaro & Barker, 2006; Burock et al., 1998; Huettel, 2012). But the increased design complexity has led to the analysis becoming further detached from the data (Huettel, 2012).

Overall, event-related fMRI has become so common that Huettel (2012) expects the term event-related fMRI to diminish as it would become an intrinsic property of functional magnetic resonance imaging (fMRI).

2.3 Face processing in humans

People encounter a variety of faces in their everyday life and can gather a lot of information from them. For the majority of humans, it takes no effort to immediately identify the face of a friend. This ability is lost in people with prosopagnosia, a face processing deficit impairing face recognition (Henson et al., 2003; Rhodes et al., 2012). Lesion studies with people with acquired prosopagnosia, as well as functional imaging, especially fMRI, studies, therefore, enable research about where face processing takes place in the human brain (Rhodes et al., 2012, Chapter 6). A "core" system of face processing was identified consisting of the occipital face area (OFA), the fusiform face area (FFA) and the superior temporal sulcus (STS) (Di Visconti Oleggio Castello et al., 2017; Rhodes et al., 2012) and is depicted in Figure 2.5. This system has a right hemispherical dominance with less consistent and weaker activation in the left hemisphere (Druzgal & D'Esposito, 2003; Rhodes et al., 2012).

2.3.1 The fusiform face area (FFA)

The fusiform face area (FFA) is located in the lateral fusiform gyrus (Collins & Olson, 2014; Rhodes et al., 2012), more specifically in its medial region (Rhodes et al., 2012, Chapter 7), and has been the focus of research in face processing (Collins & Olson, 2014). Druzgal and D'Esposito (2003) report the FFA to have an

average size of 15.3 ± 3.4 voxels. It responds stronger to face stimuli than objects (Collins & Olson, 2014; Rhodes et al., 2012). The FFA further shows higher activity after the participants report perceiving a face stimulus, even if the stimulus remains unchanged as indicated by experiments with ambiguous stimuli such as the Rubin's vase illusion (see Figure B.1) (Rhodes et al., 2012). The face representation in the FFA proved to be invariant to low-level stimulus manipulations such as scale and is suggested to be responsible for holistic face processing (Collins & Olson, 2014). The FFA seems to process structural aspects, such as spacing of facial features (Collins & Olson, 2014), which remain consistent over time (Rhodes et al., 2012, Chapter 7). Further, it shows sensitivity to angle (Collins & Olson, 2014) and identity (Collins & Olson, 2014; Rhodes et al., 2012). Upon repetition of the same face stimulus the activity in the FFA decreases.

The FFA further displays mnemonic effects: FFA activity increases with mnemonic load and is active while holding the stimulus in memory within a trial, as recorded in a delayed item recognition task by Druzgal and D'Esposito (2003).

The face exclusivity of the FFA has been questioned in research, suggesting the FFA to be an area of expertise. This would further have meant that most humans are face experts. Newer evidence, however, has disputed the expertise hypothesis (Rhodes et al., 2012, Chapter 7).

2.3.2 The occipital face area (OFA)

The occipital face area (OFA) is located before the FFA in terms of the processing order (Collins & Olson, 2014; Rhodes et al., 2012). It shows sensitivity to low level attributes such as location (Collins & Olson, 2014; Rhodes et al., 2012, Chapter 7) and to changes in a face stimulus even when those are not consciously perceived (Rhodes et al., 2012, Chapter 7). This overall sensitivity to perceptual and physical similarities suggests that the OFA plays a role in representing face parts (Collins & Olson, 2014; Rhodes et al., 2012, Chapter 7).

The OFA is functionally and structurally connected with the FFA and the STS. There is further evidence that an intact OFA is needed for the discrimination of facial identity in the FFA (Rhodes et al., 2012, Chapter 7) as it is hypothesized that the FFA integrates the from OFA represented face parts (Collins & Olson, 2014)

2.3.3 The superior temporal sulcus (STS)

The superior temporal sulcus (STS) was found to have face-selective regions, in some but not all researched subjects with occurrence rates between 50 and 75 percent. While it does not seem to be highly involved in face detection and identification its function might be in processing expression and gaze, therefore dynamic and social aspects of faces, as well as expression invariant recognition of face identity. The STS seems to remain intact in people with prosopagnosia (Rhodes et al., 2012, Chapter 7).

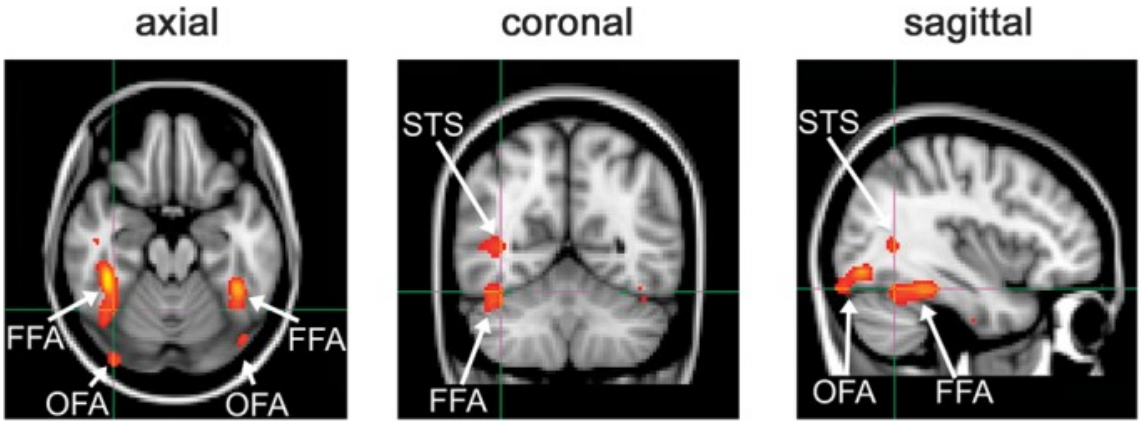


Figure 2.5: Location of face selective regions by Davies-Thompson et al. (2013). Copied from Image-Invariant Responses in Face-Selective Regions Do Not Explain the Perceptual Advantage for Familiar Face Recognition - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/Location-of-face-selective-regions-FFA-OFA-STS_fig3_221842813.

2.3.4 Face familiarity

Aspects like involvement in encoding facial expressions and the existence of familiarity effects are not resolved to date (Rhodes et al., 2012). Henson et al. (2003) found higher FFA activity for familiar versus unfamiliar faces in the lateral midfusiform, leading them to conclude that it holds a structural representation of familiar faces. Di Visconti Oleggio Castello et al. (2017) also report evidence for higher activity in the "core" face processing areas for familiar faces. They further note the robust recognition of personally familiar faces despite high disruption, while the recognition of unfamiliar faces is limited, despite high-quality images (Di Visconti Oleggio Castello et al., 2017). In contrast, Davies-Thompson et al. (2013) suggest that the representation of multiple images of a familiar or unfamiliar person in the core face-processing regions is too similar to explain the perceptional advantage of familiarity. Rhodes et al. (2012) show further examples of a lack of this familiarity effect in FFA and express that research yielded overall varying results on familiarity. There is, however, the possibility that familiarity is represented by a difference in pattern across FFA voxels or structures anterior to the FFA (Rhodes et al., 2012).

2.3.5 Face processing in the non-neurotypical population

Face processing seems to be altered not only in prosopagnosia, but also in different neurological and psychological conditions. For example, the lack of attending and using facial information is an early sign of autism spectrum disorder (ASD). However, not all people with ASD have the same extent of face processing deficits, some do not show any impairments at all (Rhodes et al., 2012, Chapter 43). The source for impairments is further hard to determine, as different processes such as atten-

tion deficits might contribute to the observed effects. Other examples of deficits in face processing, in particular facial expression labelling, have been found in major psychiatric illnesses such as schizophrenia, as well as major depressive and bipolar disorder. These deficits are disorder-specific but in all cases accompanied by functional abnormalities in regions related to facial expression perception (Rhodes et al., 2012, Chapter 44).

Despite current data supporting FFA's role in the structural encoding of faces, OFA's sensitivity to physical changes and STS's role in processing social signals, more research is needed as these conclusions remain tentative. Moreover, contrasting findings have been reported, such as in the case of familiarity processing, as well as, open questions, such as whether or not face processing is static or can adapt upon damage, remain (Rhodes et al., 2012, Chapter 7).

2.4 Artificial neural networks

2.4.1 Brief introduction to artificial neural networks

Artificial neural networks (ANNs) belong to the family of supervised machine learning techniques (Ebrahimighahnavieh et al., 2020; Yin et al., 2020), this means they learn from examples (cf. Goodfellow et al., 2016). The workings of neural networks can be summarized as follows:

> [...] the machine is given a set of labeled data, trained to discover the hidden patterns in the labeled data set, and then make predictions on the unseen data sets. (Yin et al., 2020)

In more detail, an ANN approximates a function, for example, the mapping of an input to a category (Goodfellow et al., 2016). The inputs could be images of handwritten digits and the categories the numbers 0 to 9, as in the case of the MNIST data set (see Lecun et al. (1998)). A network can learn the mapping between input and categories and thereby learn to categorize new examples. Such a network is an example of a classifier (cf. Lecun et al. (1998)).

ANNs are inspired by neural networks in the human brain (Bishop, 2006; Yin et al., 2020). Instead of interconnected neurons they consist of interconnected nodes/units, which are organized in layers and are thereby a computational graphs (cf. Goodfellow et al., 2016). A network comprises an input receiving layer, an output layer as a final layer, and hidden layers in between (Goodfellow et al., 2016, Chapter 6). The connections between these layers are uni-directional, meaning the information only flows forward through the network. This is referred to as a feed-forward network (Bishop, 2006, Chapter 5.1). Individual nodes receive weighted inputs from previous nodes, based on the sum of these inputs an activation function computes an output for this node (Bishop, 2006). This is comparable to a neuron receiving many inputs through its dendrites and if activated transmitting a signal to other neurons via the axon. The early activation functions, like the binary step function, were motivated by the all-or-non-principles of neural activation, but have since then diverged (see Goodfellow et al., 2016, Section 6.2.2.3). Especially useful are non-linear activation

functions, functions without a constant slope, as they make it possible to deal with non-linearity in the to be classified pattern (cf. Goodfellow et al., 2016, Chapter 6.2).

Based on this idea, the ANN learns as follows: Data is passed to an ANN and an output is computed. This output is then compared to its label via a loss/error function, which estimates the badness of the network (see Bishop, 2006, Chapter 5.2). In an ANN there are many different parameters, such as the weights between two connected nodes. Initially, these weights are assigned a small random value (Goodfellow et al., 2016, Chapter 6.2). To improve the network these parameters are adapted based on a mechanism called gradient descent. Gradient descent is an optimization algorithm, which finds a local minimum of the loss function (see Bishop, 2006, Section 5.2.1 and 5.2.4). If the loss function is minimized the performance of an ANN is maximized. To minimize the loss function, error information must travel through the network, which is realized by an algorithm called backpropagation (Bishop, 2006; Goodfellow et al., 2016, Chapter 5.3, Chapter 6.5). Together, these two algorithms make the training of a network possible. Different approaches to gradient descent are used, which differ in the amount of training data considered for each parameter update. Often stochastic gradient descent, that works of one or a few samples at a time, is used for classical ANN training (see Lecun et al., 1998, Goodfellow et al., 2016, Chapter 5.9). The learning rate determines how much the parameters are updated each time by the new information from gradient descent, with a higher learning rate meaning the parameters are updated to a higher extend (cf. Bishop, 2006, Section 5.2.4).

Convolutional neural networks

Convolutional neural networks (CNNs) are ANNs commonly used for image analysis. Their workings are inspired by the functionality of the visual cortex (Ebrahimighahnavieh et al., 2020). A CNN tries to identify local features of an input image, such as horizontal or vertical lines. Layers deeper in the network can extract higher-level features, for example edges, by merging lower-level features (Bishop, 2006, Section 5.5.6). Instead of having weighted links between nodes, like in the classical ANN, a CNN has weight matrices, so-called kernels (Yin et al., 2020), which traverse over an input. One could imagine these kernels to represent some sort of feature, for example, a horizontal line, which then slides over an image and looks for this feature locally (see Bishop, 2006, Section 5.5.6). If one takes a look at the MNIST data set and compares the self-written digits one notices similarities between digits of the same class. For example, the attributes which make a handwritten Arabic numeral four a four are local features like specific edges that cannot be found in other numbers. A kernel representing a horizontal line would be able to detect the horizontal line discernible in the middle of a four. The result of kernels traversing over the input are feature maps, that can serve as an input to a pooling layer, which reduces the size of feature maps by merging operations, or another convolutional layer (Bishop, 2006, Section 5.5.6). The benefit of this method is that the CNN is insensitive to shift and scale of the input (Bishop, 2006, Chapter 5), as the kernels are able to find a feature independent from its position or size. Further, compared

to other networks of the same size, CNNs are easier to train and require less data for a good performance due to them having fewer parameters (Yin et al., 2020).

2.4.2 Artificial neural network applications in neuroimaging

There has been a recent interest to use machine learning methods with inputs from neuroimaging (Zhao & Zhao, 2020). Artificial neural networks have been successfully applied to various imaging modalities, such as EEG (Schirrmeister et al., 2017), MEG (Zubarev et al., 2019), MRI (Zhao & Zhao, 2020) and fMRI (Vu et al., 2020). Artificial neural networks have been successfully used to classify neurological disorders and psychological diseases. These include, but are not limited to, attention deficit hyperactivity disorder (Riaz et al., 2020), schizophrenia (Yin et al., 2020), Alzheimer's disease (Ebrahimighahnavieh et al., 2020) and autism spectrum disorders (ASD) (Yin et al., 2020). Their success shows the potential of neural networks to improve diagnoses and possibly becoming the de facto solution for imaging problems in the medical field (Yin et al., 2020). Since ANNs are only able to do classification in regard to one disorder at a time so far, it would be of immense benefit to expand that scope (Yin et al., 2020). Besides disorder classification, CNNs have been used to classify experimental tasks (Vu et al., 2020), as well as to predict the BOLD response to an image stimulus (Zhang et al., 2019). It is important consider the networks interpretability for the development of diagnosing ANN as the identification of a specific biomarker is clinically crucial (Yin et al., 2020).

2.4.3 Vu et al. (2020): fMRI volume classification using a 3D convolutional neural network

Personal difference in brain architecture, such as differences in the location of the FFA, exist. To classify single brain volumes from different participants nonetheless correctly it is important that a machine learning method trained on such data can deal with this interpersonal variance. This is why the CNN presented by Vu et al. (2020) was chosen for the study presented in this thesis.

Vu et al. (2020) propose a convolutional neural network for the classification of task-based fMRI volumes. They hypothesized that such a network would not be affected by shifting and scaling in neuronal activation and thereby be able to deal with spatial misalignment which cannot be corrected during preprocessing of the data.

Their data stems from twelve healthy volunteers engaging in a sensorimotor block design, in which they were tasked with clenching their left hand, as well as their right hand and attending auditory as well as a visual stimulus. A 3T MRI scanner with a two-second TR and 30ms TE was used. This way, 120 volumes over four tasks per person, with a total of 1440 volumes for all participants, were acquired. The data was preprocessed to different extents using SPM8, including some or all of the following steps: slice timing correction, motion correction, MNI spatial normalization and smoothing with an 8-mm full width half-maximum Gaussian kernel. Afterwards, a brain mask was applied, such that only signal within the brain is considered.

Then the volumes were normalized to percent signal change (PSC) using the resting scans.

The CNN by Vu et al. (2020) is a modification of LeNET5 and consists of three convolutional layers which all have the same stride size. The first layer has 8 filters, with a 7x7x7 kernel. The second layer consists of 16 filters, with a kernel size of 5x5x5 and the third convolutional layer has 32 filters with a kernel size of 3x3x3. Following the convolutional layer are two fully connected layers. The training algorithm used is stochastic gradient descent with cross-entropy loss and a mini-batch size of 50. The CNN was trained for 50 epochs, meaning the full training data set was passed through the network 50 times. They annealed their learning rate starting from 10^{-3} leading to 10^{-6}. They further tested different cross-validation schemes with leave one out cross-validation performing the best.

Vu et al. (2020) compared their results to a 1 dimensional fully connected ANN as well as other classifiers. They found their CNN to have lower error rates compared to both the ANN and the other classifiers. They further found the error to vary significantly between subjects. Using visualization techniques, such as feature and saliency maps, they could investigate which parts of the fMRI input the CNN used in decision making. In this way they were able to make inferences on which brain parts were relevant for the classification. It was found that the feature maps, though shifted between the individual convolutional layers, showed task-specific brain areas, thereby revealing a CNN based biomarker. They were further able to test the CNN with alternative data from the Human Connectome Project (van Essen et al., 2012), as well as in online classification with both showing good results.

2.5 Related work

Different studies have worked with event-related fMRI data for classification purposes (Yin et al., 2020). Including Mumford et al. (2012) who utilized regression coefficient from a general linear model (GLM). Such models are classically used for single subject statistical analyses by software packages such as SPM (see The SPM Developers (2020)). Others created an encoding model before decoding (Zhang et al., 2019). Also, the usage of multivariate Bayesian models was proposed (Lee et al., 2017).

3 This Study

3.1 Research Questions and Focus

The focus of this study lies in finding a processing pipeline enabling classification of full, 3-dimensional, fMRI volumes, acquired from a face viewing paradigm. The goal of this thesis is to answer the following research questions

1. Is it possible to create data-label pairs from (task-based) fast event-related fMRI data for artificial neural network applications?
   - 1a. Can this be achieved without using a general linear model?
   - 1b. Can this be done by changing the trial definition/ the trial labelling?
     - i. Is there an optimum to the trial labelling?
     - ii. Is this optimum global or local?
2. How is the classification performance of a CNN for a task-based event-related fMRI input?
   - 2a. Does the CNN utilize known neurological correlates related to the experimental paradigm when classifying fast event-related fMRI?
   - 2b. Is there a classification difference between the optimal trial labelling and a not optimal trial labelling?

This study uses fast event-related fMRI data from Henson et al. (2003). It will process them, such that they can be used as input for the CNN proposed by Vu et al. (2020).

3.2 Materials and Methods

3.2.1 The data set

The open-source data set by Henson et al. (2003) is multi-modal and consists of EEG, MEG, diffusion-weighted MRI, MRI and fMRI data. This study only uses the fMRI and MRI subset. The MRIs were recorded on a Siemens 3T Trio: The MRI are 1 mm isotropic T1 MPRAGE images (TR: 2,250ms TE: 2,98), the fMRIs are axial, echo-planar imagining sequences (TR: 2s, TE: 30, interleaved slice acquisition). The fMRI data set includes the data of 16 young healthy adult participants. The experimental stimuli set used consists of 300 greyscale images of famous, unfamiliar, and scrambled faces. The faces had mainly a happy or neutral expression and were taken from 3/4 or full-frontal perspectives. To remain attentive, the participants

were tasked to judge the symmetry of the presented stimuli. Due to the subjectivity of this task, this behavioural data is not of interest to the present study.

The experiment employed a fast event-related design using jittering and was structured as follows: A fixation cross appeared for 400-600ms followed by the face or scrambled stimuli, which were presented for 800-1000ms. Afterwards, a circle was presented for 1.7 seconds before the next trial began. After 50 seconds of task trials, a 20 second rest period was integrated. Nine runs consisting of about 30 trials for each condition (famous, unfamiliar and scrambled) were carried out in this way. Each of the nine runs led to 208 volumes (the first two scans were removed), with the exemption of the last run of participant 10, which comprises only 170 volumes due to a scanning error. All volumes were anonymized by removing the face. The DICOM images were transformed to NIfTI format.

The entire data set was downloaded from Rik Hensons laboratory website[1].

Trial definitions

The trial definitions included in the downloaded data set missed information about the duration of the subparts of the trial and only included the time of trial onset. This is why an additional, more detailed, trial description was downloaded from the openneuro data sets repository [2]. This trial definition holds additional information including the start of the fixation period and the stimuli presentation, as well as the onset of the rest periods.

3.2.2 Computation

Aside from the CNN, all code was written and executed with MATLAB R2019a. All scripts were first drafted and tested with a single hardware system (Intel(R) Core(TM) i5-7200 CPU 2.50GHz, NVIDIA GeForce 940MX, 8GB RAM). All final results were computed on the Cedar division of Compute Canada[3], a heterogeneous computing cluster, allocating varying resources. To improve execution time all MATLAB scripts with high amounts of read and write operations were adapted such that they make use of the temporary storage on the RAM disk (refer to the Compute Canada wiki on handling large collections of files [4]). The scripts used on Cedar as well as the single hardware system were uploaded to GitHub[5].

[1] Downloaded between 12.08.2020 and 20.08.2020 from ftp://ftp.mrc-cbu.cam.ac.uk/personal/rik.henson/wakemandg_hensonrn/. As of March 2021 this link is no longer functional. An alternative location for the data is https://openneuro.org/datasets/ds000117/versions/1.0.3

[2] Downloaded on the 29.09.2020 from https://openneuro.org/datasets/ds000117/versions/1.0.3

[3] See www.computecanada.ca

[4] https://docs.computecanada.ca/wiki/Handling_large_collections_of_files for further information

[5] See https://github.com/htscode/bachelor_thesis_2021

3.2.3 Preprocessing

The preprocessing of the data was carried out using SPM12 (7771) and is based on chapter 42 of the SPM Manual (The SPM Developers, 2020). The processing steps described in the manual are mostly in line with the preprocessing used for artificial neural network applications (Ebrahimighahnavieh et al., 2020; Lee et al., 2017; Vu et al., 2020), however, lack the often used step of slice time alignment. Therefore, this step was added as the first step in this studies preprocessing. The preprocessing steps now match those used by Vu et al. (2020) and are as follows:

Since individual slices of the brain were recorded one after the other over the whole trial repetition time (TR), which is 2 seconds for this data, they display vastly different time points. Slice timing correction corrects for this so that all slices of one volume appear to have the same acquisition time. This data was aligned to the first (lowest) axial slice.

Slice timing correction was followed by the step of realignment, in which the individual fMRI scans were spatially aligned to one another, so that movement artifacts, created by translations and rotations of the head, were reduced.

Afterwards, the normalization parameters were estimated. This creates a deformation field, which was applied to the images later to warp them to MNI152 space. Such an alignment to a common space makes it possible to compare results between individuals and studies. Note that only the structural image was used for the estimation of a deformation field.

This process was followed by coregistration. This step was a preparation for the application of the previously identified normalization parameters. In order to apply the normalization to all fMRI images they needed to be coregistered to the structural image, for which the deformation field was calculated. In detail, SPM tried to find a transformation that describes how each individual scan relates to the structural image. Thereafter, the deformation field was applied, thereby finalizing the normalization step.

The last step in preprocessing was the application of an 8mm isotropic Gaussian kernel to smooth the images. In this way, artifacts were suppressed and the probability that the activities of multiple participants have spatial intersections in the data increased. The smoothing can be seen in the scans as a blurring effect.

More details about the individual steps and relevant variables can be found in the SPM12 manual (cf. The SPM Developers, 2020), as well as in the batch editor of SPM12.

3.2.4 The time components of fast event-related fMRI

In order label the fast event-related data, multiple time components had to be considered. These components are exemplified in Figure 3.1 where a), b) and c) are different time components, with an additional time axis in seconds being displayed for orientation.

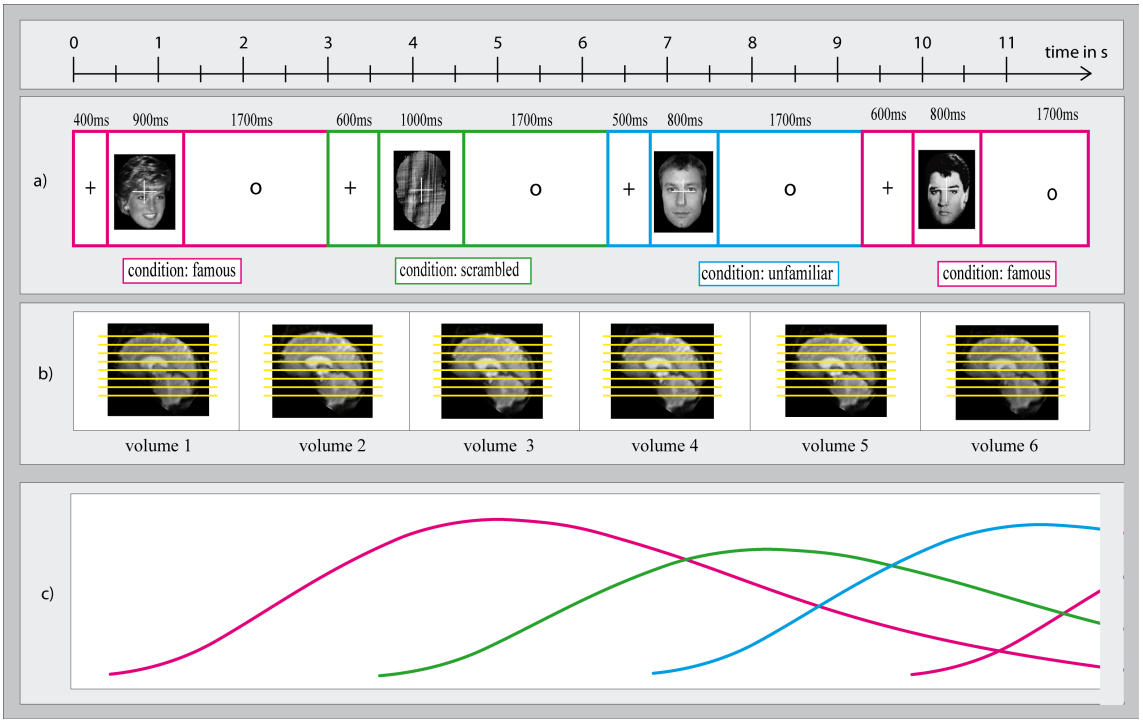


Figure 3.1: Schematic drawing of the time components of fast event-related fMRI, a) *trial definition*: + marks the fixation period, the *image* the stimuli, and the *o* the inter-trial period, the duration of each is indicated above them. Magenta, green and blue indicate a famous, scrambled or unfamiliar trial, respectively. b) *scan time*: with a scan duration (TR) of 2 seconds (length of white box) and yellow lines illustrating the axial slice acquisition c) *hemodynamic response*: simplified approximation of the hemodynamic response function in face selective regions.

The first time component of interest is the *trial definition* or *trial timing*, part a) of Figure 3.1. This component describes how and when the trials were presented, which includes the onset of a trial, as well as the length of the fixation period, how long one stimulus was presented and the onset and duration of the rest period. Due to the usage of jittering, a new trial can begin between after 2.9 to 3.3 seconds.

The second time component is the *scan time* as shown in part b) of Figure 3.1. To complete one whole brain scan the scanner used by Henson et al. (2003) took two seconds. Recall that one brain volume consist of many, in this case axial, slices, which are captured individually. Further note that the scan start is independent of the trial onsets.

The third time component, c) in Figure 3.1, is the *hemodynamic response*. As discussed in section 2.1, the hemodynamic response is delayed, meaning the response to a stimulus at time point 0 is the highest $3 - 9$ seconds later due to the BOLD signal delay. In the example in Figure 3.1, this leads to the peak of the BOLD response for the first famous trial occurring at around five seconds and, therefore, a BOLD signal delay of roughly four seconds. At this time volume three is acquired and a scrambled trial presented. Consequently, volume three does not represent the

neural response to the then presented scrambled stimulus, but rather that of the famous stimulus from the trial before.

This misalignment of these three time components makes labelling of the brain volumes difficult. The following problems arise: Firstly, "overlapping": Due to the characteristics of the hemodynamic response in a fast event-related design responses to different stimuli overlap (see subsection 2.2.2). This means that at some points in time the BOLD signal comprises the response to more than one stimulus. In Figure 3.1 this can be observed when looking at the BOLD response between 6 and 7 seconds, where the response to the first famous stimulus and the scrambled stimulus overlap significantly. The brain scan which was created at this point in time, in Figure 3.1 volume 4, reflects the neural response to two stimuli, which is problematic for labelling purposes, as no definite label can be assigned to such a volume. Even though there is overlap at other points in time, e.g. at 5 seconds, the estimated response to one stimulus is significantly greater, consequently, dominating the signal.

Secondly, "Delay": the scan taken during the presentation of a trial does not reflect the maximal response to the stimulus presented in the trial due to the BOLD signal delay.

Thirdly, "scan trial mismatch", due to the misalignment of the trial timeline and the scan timeline it is often the case that a trial onset is in the middle of a brain scan. This makes it harder to label this scan as parts of two different trials were presented within one scan.

The first and third problem are similar since both lead to difficulties in assigning a definite label to a scan. However, the time components responsible for this are different. The "overlapping" is a problem of the misalignment of the second and third time component, while the "scan trial mismatch" is a problem of the misalignment of the first and second time component.

The label given to a brain scan should reflect the BOLD response rather than the trial presented at the acquisition time, therefore rendering the third problem of "scan trial mismatch" irrelevant. However, the "scan trial mismatch" was used as an information component to solve the "overlapping" problem in this study.

The only clue for labelling is given by the condition of a trial. However, due to the misalignment of the time components, the condition of the trial presented at a time point t cannot be confidently used as a label for the scan acquired at the same time point.

In order to solve the problems created through the misalignment of the components, this study realigned the components, such that a volume's label matches the stimulus response it captures. As the time components' scan time b) and hemodynamic response c) are fixed to each other and cannot be adjusted post acquisition, component a), the trial definition, was changed. This change can be described as a shift of the trial definition forward such that it might align more closely with the BOLD response. This was realized by two algorithms: a shifting algorithm, which shifts the trial definition forward, and an optimization algorithm, which optimizes this time

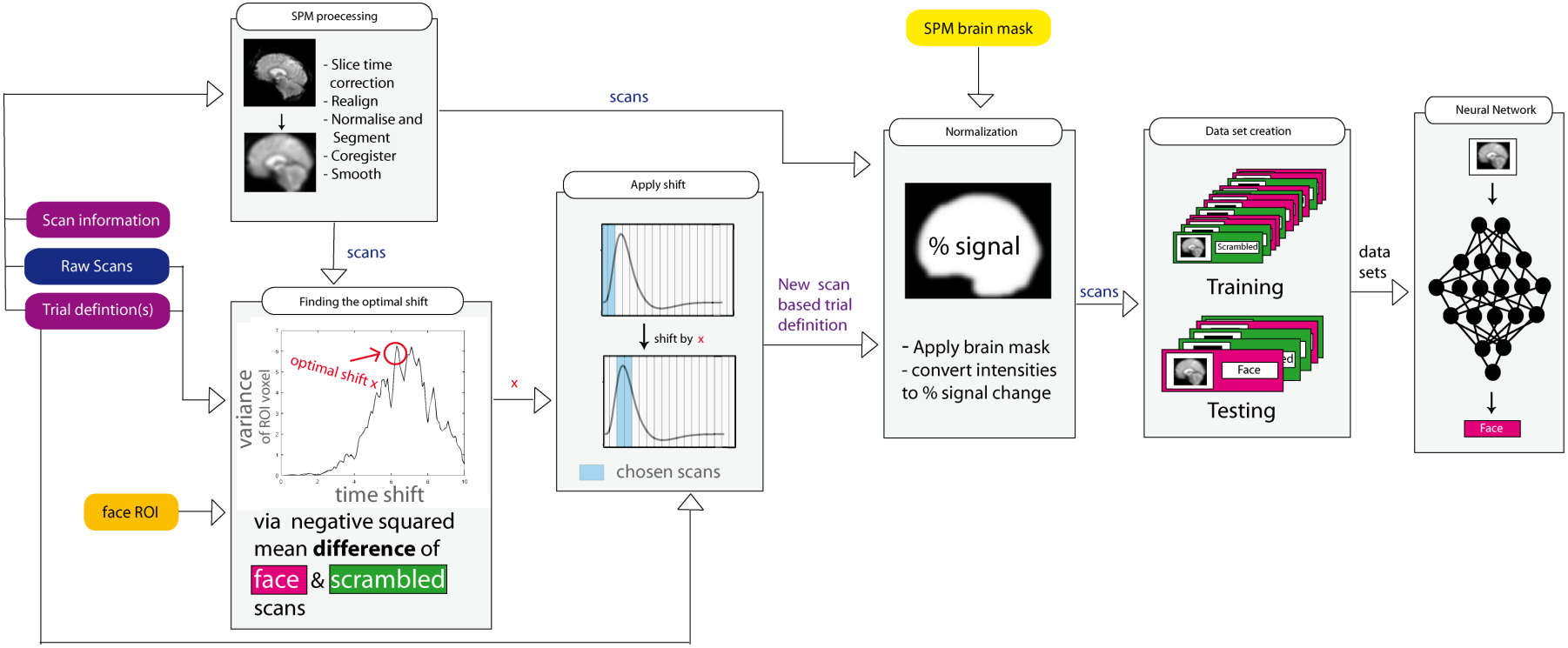


Figure 3.2: Processing pipeline consisting of the steps: Preprocessing, optimization, applying shift, normalization, data set creation and applying a neural network.

shift.

This approach resulted in the processing pipeline seen in Figure 3.2 and consists of the steps preprocessing, optimization, application of the best shift, normalization, data set creation and classification with the CNN of Vu et al. (2020).

3.2.5 Optimization

It is unknown what the best time shift is to align the time components and thereby incorporate the BOLD signal delay, such that a scan labelled scrambled also holds the peak response to a scrambled stimulus. To find this time shift optimization was used. From literature, it is known that between face stimuli and scrambled stimuli a difference in BOLD signal can be found (see section 2.3).

The optimization algorithm uses these differences, as described in Equation 3.1, by calculating the mean of all scrambled scans and deducting it from the mean of all face, i.e all famous and unfamiliar, scans. More precisely the NSMD is calculated for every voxel individually. To accentuate even small differences, the results were squared. As further optimization algorithms, including MATLABs optimization techniques, operate as minimizers by default was the squared mean difference multiplied by minus one. Therefore, the algorithm could be used to find a minimum by calculating the NSMD based on different trial definitions stemming from different time shifts. The trial definition of the time shift leading to the biggest negative squared mean difference (NSMD) was then deemed the most optimal trial definition, and thereby the best scan labelling.

$$-(\sum facescans_{voxel(s)} - \sum scrambledscans_{voxel(s)})^2 \tag{3.1}$$

Based on Logothetis and Wandell (2004) and other work on BOLD signal delay, a lower boundary of 0 and an upper boundary of 10 were determined for the time shift.

Using MATLAB optimization techniques

The first approach was to use classical optimization algorithms. For this, only derivative-free methods are applicable as the underlying function is unknown and an algorithm is in place of a function. Downhill simplex and simulated annealing were tested[6].

Simulated annealing with the standard stop criteria was not able to terminate, possibly due to jumping from one local minimum to another as a result of too strict stopping criteria. Preliminary results were calculated by stopping calculations after 8 hours and reporting the best minimum which was found until then.

[6]For both just one voxel was considered for the NSMD calculation as the minimization of just one variable is performed

The application of downhill simplex was also unsuccessful. All results were worse than the best results achieved by simulated annealing. Thereby, it was concluded that the underlying function is not smooth.

Following this exploration, it was decided that the knowledge about the underlying function, mapping the negative squared mean difference (NSMD) upon time shift, seems to be insufficient to confidently apply classic optimization techniques. Therefore, a so-called brute force method was applied, presupposing to, thereby, also get more information about the underlying function.

Brute force method

A brute force method relies on trying many possibilities until a desirable result is achieved (Colman, 2009). In this case, over 100 different time shifts were tried and their results were saved. The method was applied to every participant individually as the BOLD signal can differ among individuals (Logothetis & Wandell, 2004). In particular, every shift from 0 to 10 at a scale of 0.1 seconds was tried, leading to 101 samples. The brute force method is further described in algorithm 1.

Three implementations of this algorithm were written, differing in the number of voxels used for the NSMD calculation. Unlike the other optimization methods, the brute force method is not dependent on having a single variable that is minimized, making an approach with more voxels possible. The "all voxel" version calculates the NSMD for all brain voxels. This, however, leads to unpractically large samples with more than half a million entries. As not all voxels correspond to brain regions important in face processing this approach was dropped.

An alternative was the usage of one specified voxel in the FFA. However, due to the interpersonal difference in the location of the FFA among participants this approach failed for some of the participants. This led to the development of an algorithm to improve the choice of the FFA voxel. As this approach was dropped in favour of a third approach this voxel finding algorithm was not of any further use.

The third approach was the usage of a region of interest (ROI) related to face activity.

Algorithm 1 Brute force algorithm for ROI

procedure BRUTE FORCE
    $ROI \leftarrow$ region of interest
    $shifts \leftarrow [0.0, 0.1, ..., 10]$
    $scans \leftarrow$ all fMRI scans
    $trialDef \leftarrow$ trial definition of all runs
    for i = 1:length($shifts$) **do**
        $newTrialDef \leftarrow$ shiftingAlgorithm($shifts$(i), $trialDef$)
        $facescans \leftarrow$ all famous or unfamiliar $scans$ **in** $newTrialDef$
        $scrambledcans \leftarrow$ all scrambled $scans$ **in** $newTrialDef$
        $voxels \leftarrow$ all voxels of ROI
        $NSMD(i) \leftarrow -(\sum facescans_{voxels} - \sum scrambledscans_{voxels})^2$
    end for
    $[-, bestShift] \leftarrow max(variance(NSMD'))$
end procedure

Choosing a ROI

Despite its important role in face processing the FFA was not chosen by default as a ROI, because it is not certain that FFA voxels work superior to other face processing regions when employed in the brute force method. Instead, the website neurosynth was used to specify a region of interest [7]. Neurosynth provides an automatic meta-analysis of neuroimaging studies. It then creates a NIfTI image which represents the activation's found in the different studies (Yarkoni et al., 2011). In this way a specified ROI does not rely on one particular study but rather on the combined knowledge created through different studies.

A term-based meta-analysis with the term "face" was used, which takes 896 face-related studies into account. As recommended by the website the "association test map" was employed (cf. Yarkoni et al., 2011). To reduce the computational load of the brute force algorithm, only the strongest 0.01% of all voxels were used. In order to exclude higher brain functions and emotions, a primitive exclusion for related areas was added by excluding all voxels with a y-coordinate greater than -15. The remaining 226 voxels corresponded mostly to FFA regions, with a few voxels in the OFA. The final region of interest is visualized in Figure 3.3.

[7] See https://neurosynth.org/

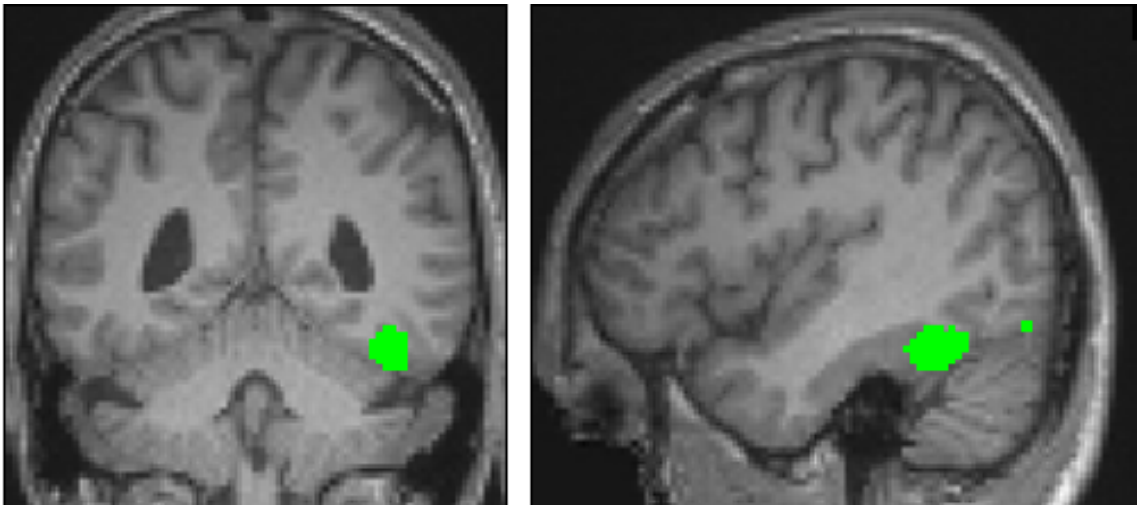


Figure 3.3: ROI overlayed onto the normalized structural image of participant 15, coronal and sagittal view, right hemisphere.

Shifting algorithm

The optimization algorithm uses for the calculation of the NSMD at all time shifts a time shift specific trial definition which is the results of the here presented shifting algorithm. The shifting algorithm creates a new trial definition based on the existing one, this new trial definition leads to a better alignment of the time components making scan labelling based on the trial definition possible. This new trial definition is scan-based, meaning that every scan is seen as an individual trial of one of the three conditions: famous, unfamiliar and scrambled, which further automatically serve as a scan's label. Scans to which no definite label could be assigned were excluded. For this exclusion, information gathered from the "scan trial mismatch" was used.

The algorithm requires a time shift as input and works as follows: All trials get shifted by an individual shift, which is the sum of the input shift and the fixation period. If no shift to correct for the BOLD signal delay is desired, the input shift should be 0 and the individual shift equals the fixation period. To make each scan into an individual trial, the algorithm looks at the scans happening during a shifted trial. All scans happening during such a trial are each a new trial of the same condition as long as the volume acquisition was finished before the next shifted trial's onset. A scan which started in one trial but will end during another trial is excluded unless both trials are of the same condition or a rest period is in place of the second trial.

The algorithm considers two special cases. Firstly, scans which started before the first trial after the shifting of the trials are excluded. For example, if the time shift was 6 seconds, the first three scans are excluded and the first shifted trial onset is at 6 seconds instead of zero. Secondly, due to the last trial being neither a rest period as defined by Henson et al. (2003) nor a famous, unfamiliar or scrambled trial all scans of this trial are excluded.

For all participants and for each of the nine runs a new scan based and shifted trial definition is created in this way for each time shift. To find the best time shift a new temporary trial definition gets created within the optimization algorithm for every time shift.

In order to know when a rest period occurs a second version of the shifting algorithm was programmed, which also includes rest as a condition. This version was utilized for the normalization purposes explained in section 3.2.6.

Selecting the best shift

In order to select the best shift based on all calculated NSMDs, the space variance was utilized (see algorithm 1 operation 13). This approach was borrowed from MEGs and EEGs global field power. Global field power is a method used to quantify the amount of activity in the M/EEG field at a time point while avoiding a domination of near zero activity (cf. Skrandies, 1990). In this study the space variance was calculated over time shifts rather than time points. However, it served the same purpose of avoiding that voxels with little activity dominate the results and, thereby, the choice of the best shift. Note that due to the variance being always positive, a maximum ends up being selected.

3.2.6 The neural network

The python implementation of the CNN by Vu et al. (2020) was used, which was realized in a jupyter notebook and can be found on GitHub[8]. It shall be noted that this network was implemented in the first version of TensorFlow. The same version was used in this study, even though it is deprecated. Due to the large data set and the limited memory capabilities of the utilized GPUs only a third of all data could be used without modifying the network. Consequently, the first three runs of each participant comprised the data set. One participant was randomly chosen to be the test subject, this was participant 15.

At first the CNN was trained for 50 and 10 epochs on the data set resulting from the new trial definition based on the optimal shift found via brute force optimization. This optimized shift data set had a training set size of 4645 scans and a testing set of size 317 scans. The second time the CNN was trained on a data set that resulted from using a non optimal, or more specifically zero shift, for the same number of epochs. This data set had a similarly large training set of 4630 samples and a testing set 316 with scans. The CNNs were trained via a Python script converted from the JuPyter notebook to permit the submission of the script as a batch job on the computing cluster. In all cases, Python version 3.7 was used.

[8]See https://github.com/bsplku/3dcnn4fmri/tree/master/Python_code

Voxel signal normalization

Vu et al. (2020) used normalization techniques to further process their data before using it as a neural network input. For this, they applied a brain mask and then calculated percent signal change (PSC)s for every voxel. Alternatively to Vu et al. (2020), who used an activation based brain mask, this study utilized the brain mask of the SPM12 software package. The anatomical brain mask is applicable here, as it sits in MNI152 space and is expected to include more brain areas than the mask by Vu et al. (2020). The SPM brain mask was scaled to the functional images and converted to binary via the SPM12 GUI. With the brain mask all voxels which were not in the brain could be found and set to 0.

Afterwards, the PSCs were computed with the mean rest periods as the baseline. This converted the unit at each voxel from intensities to a more comparable and interpretable format. The resulting PSC is the activity change the presentation of stimuli lead to compared to the resting baseline as presented in Equation 3.2. Due to the Henson et al. (2003) data consisting of multiple runs, a mean rest baseline was computed for each run individually. As no formula was provided by Vu et al. (2020) a classical percent change computation was used.

$$\frac{v_{xyz} - \overline{rv_{xyz}}}{\overline{rv_{xyz}}} \cdot 100 \tag{3.2}$$

Where v_{xyz} denotes an individual voxel and $\overline{rv_{xyz}}$ denotes the corresponding mean rest voxel.

Afterwards, the final data set was created. At this step all unfamiliar and famous scans received the label "face" corresponding to 1, all scrambled scans received the label "scrambled" and the binary 0 correspondingly. The normalization was done for both the data with a label according to the optimal shift and the data with a label corresponding to a shift of zero.

4 Results

4.1 Optimization results

The execution time for the brute force algorithm was around 1 hour and 50 minutes varying per execution while using two CPUs and 16 GB RAM with a batch script on Cedar.

In Figure 4.1 the results of the brute force method for the data of participant 15 are shown. Every line corresponds to one voxel, with the y-axis describing the negative squared mean difference and the x-axis the time shift[1]. A more negative value corresponds to a greater difference between the scrambled and face conditions. Overall, a bell shaped form or dip with a minimum at around 6 seconds is visible. Some of the voxels reside around 0, displaying a very low range, while a few voxels have a more extreme range. Further, voxels of high range block together with a decreasing range with greater distance (see Appendix Figure B.2). The range herein describes the spread of the NSMD in one voxel, a high range corresponds with a great NSMD.

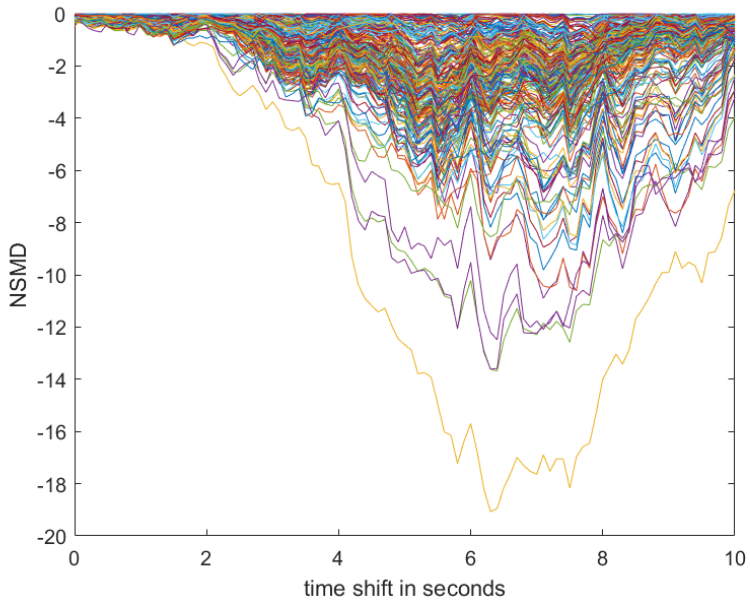
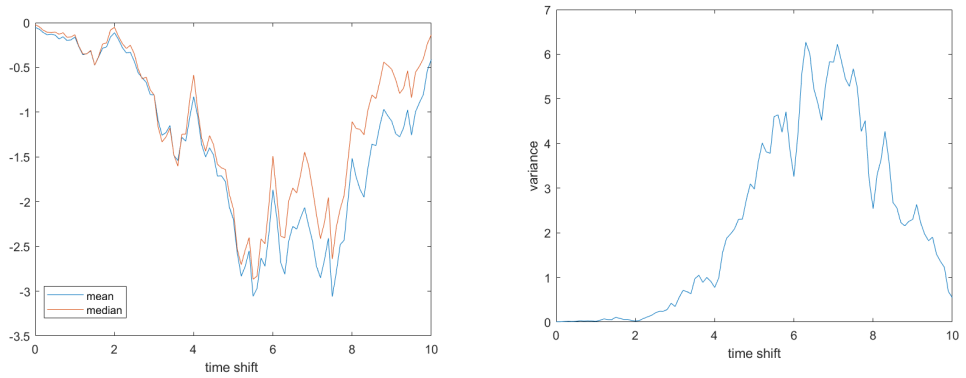


Figure 4.1: Brute force results: time shift to negative squared mean difference of face and scrambled condition, participant 15.

Figure 4.2 a) shows the mean and median of all voxels, two potential minimums

[1]The colour has no significant meaning besides every voxel having its own colour

are visible at 5.5 and 7.5 seconds. The variance, depicted in Figure 4.2 b) reveals a roughly bell shape curve with the highest peak at 6.4 seconds, which was consequently chosen as the optimal shift for participant 15.

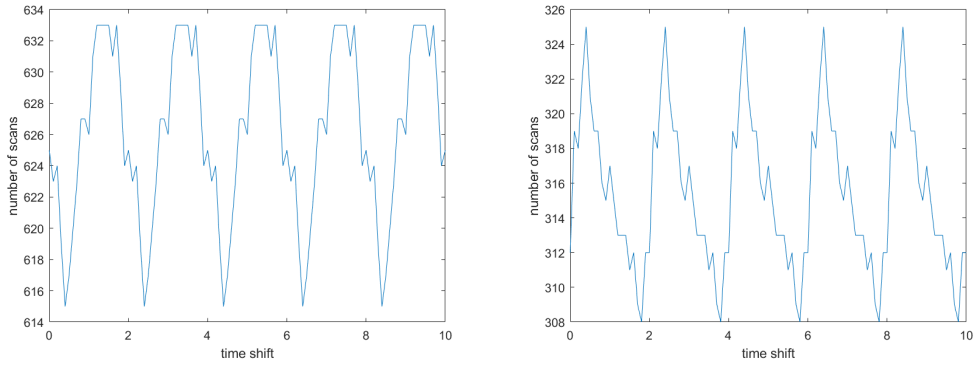


(a) Mean and median of all voxels' negative squared mean difference.

(b) Variance of all voxels' negative squared mean difference.

Figure 4.2: Brute force results participant 15.

When looking at the number of trials taken into account at every time shift, as plotted in Figure 4.3 for participant 15, a repeating pattern is visible. Every two seconds the pattern repeats, which coincides with the two-second TR. The number of trials differ only slightly, in case of participant 15 by less than 3 % or just over 5%, when comparing the highest number and lowest number of trials considered for the face and scrambled condition respectively.



(a) Number of scans used per time shift in the face condition.

(b) Number of scans used per time shift in the scrambled condition.

Figure 4.3: Brute force results: number of scans over time shift a) face condition, b) scrambled condition.

The optimal shift resulting from the application of the brute force method for all participants can be found in Table 4.1. The optimal shifts according to the variance lay between 3.7 and 6.4 seconds, while 12 out of 15 are in the range of 4.4 and

5 seconds. The median and mean show greater variance in the results, with the median depicting results that are biologically infeasible. In particular the median of all voxels' NMSD of participants 5 and 6 is at over 9 seconds, a labelling according to this time shift can not reflect the peak of the hemodynamic response.

Table 4.1: Optimal shift per participant using mean, median or variance.

Participant	**Mean**	**Median**	**Variance**
1	4.5	4.5	4.9
2	3.3	3.3	4.4
3	5	5.3	5
4	3.7	3.7	3.7
5	5	9.9	5
6	3.8	9.4	4.8
7	4.6	4.6	4.6
8	4.6	4.6	4.6
9	4.9	4.8	4.9
10	4.6	4.6	4.8
11	4.8	6.5	4.7
12	4.3	4.4	4.2
13	5	5	5
14	6.1	6.1	4.8
15	7.6	5.6	6.4
16	5.5	4.3	4.9

Applying the optimal shift as input to the shifting algorithm leads to a new trial definition. It is possible to compare this trial definition with a trial definition computed from the shifting algorithm with 0 as shift input. Note that even with no shift, exclusion of scans via the exclusion criteria and shifting related to the fixation period is still applied. With the help of an alluvial flow diagram[2] the difference in the trial definition resulting from a 0 and optimal shift for a participant can be visualized. The alluvial plot for the condition label of participant 15 is shown in Figure 4.4.

[2]Refer to https://de.mathworks.com/matlabcentral/fileexchange/66746-alluvial-flow-diagram.

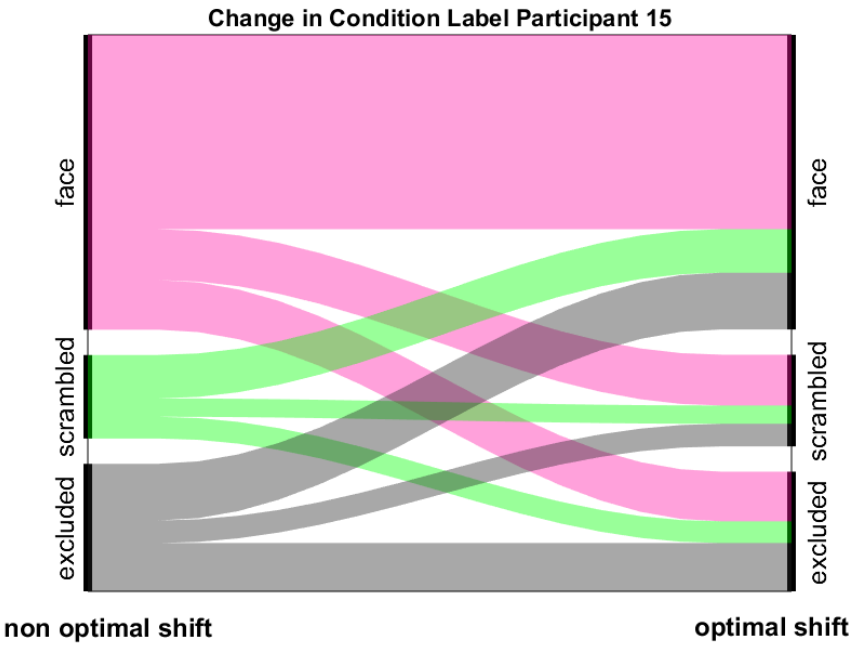


Figure 4.4: Alluvial flow diagram representing the change in condition label comparing a shift of zero (left) to the optimal shift (right),which lies at 6.4 seconds, for participant 15.

How large the overlap of the data set with an optimized shift (OS data) and a not optimal shift (NOS data) is can be assessed by examining whether or not the same scans are present in both data sets. The overall overlap of the two data sets lies at 64.96%, meaning that most scans are present in both data sets. Further comparing the overlap of the face condition reveals that less than half (43.52 %) of the scans are considered a face trial in both data sets. For the scrambled condition, the overlap is even lower such that for the scrambled condition less than a third (21.01 %) of the scans are present in both data sets. Further, 19.53 % of those scans which have been labelled scrambled in the NOS data are labelled face in the OS data. Of those scans labelled face in the NOS data 23.53 % are labelled scrambled in the OS data.

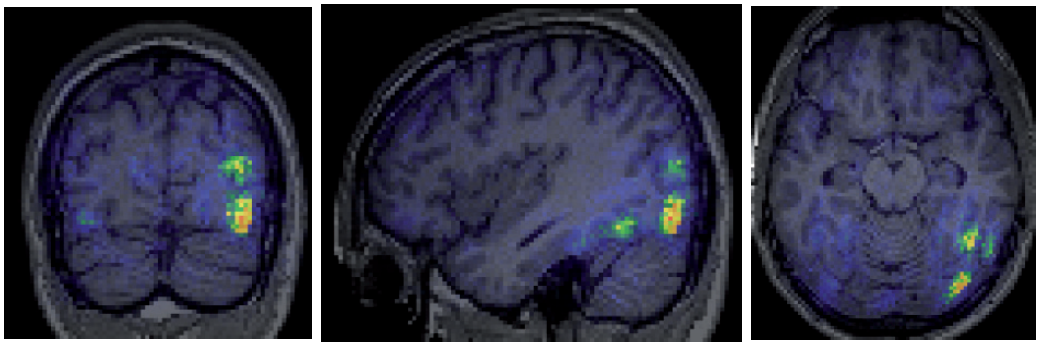
4.2 The performance of the artificial neural network

The CNN trained on the data with an optimized shift (OS data) over 50 epochs has a 99.37 % test accuracy, with two false-positive scans (classified as face condition but have the label scrambled). The network trained with a zero shift (non OS data) achieved a 100 % accuracy when training for the whole 50 epochs. When trained for only 10 epochs the OS data achieved an accuracy of 91.80 %, while the non OS data achieved an accuracy of 77.85 % .

The training time for the OS data with 50 epochs was 3 hours and 34 minutes, for the non OS data was 2 hours and 54 minutes, respectively. Both of the 10 epoch results were gathered in an interactive Jupyter Notebook session not allowing for

correct time measures.

The saliency visualization technique from Vu et al. (2020) was used to make inferences about the CNNs feature representation. All saliency maps obtained from individual trials from the test subject were averaged and plotted on the respective MRI volume[3]. A high saliency corresponds to an area being of high importance for decision making for the CNN. Due to the classification being binary, the saliency maps of both the face and scrambled condition are very similar. The saliency maps for the scrambled conditions can be found in the appendix (see Appendix A). The results for the optimal shift data set for the face condition are shown in Figure 4.5. It is recognizable that face-selective regions in the right hemisphere show high saliency, namely FFA and OFA regions. There is no comparable saliency in STS regions. Further, there seems to be a slightly heightened saliency for left hemispherical OFA regions.

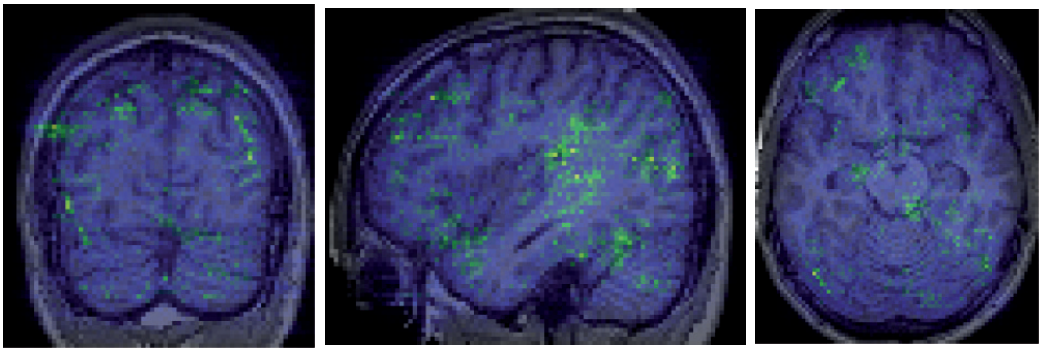


(a) Coronal slice (b) Sagittal slice (c) Axial slice

Figure 4.5: Saliency map from the OS data set for participant 15 face condition, coronal, sagittal and axial slice at 40x-82x-16mm.

The non optimal shift data set led to the saliency map in Figure 4.6. The saliency is overall more scattered across the brain with a slightly elevated saliency in right hemispherical FFA regions, no higher saliency in OFA regions and higher saliency blobs around the caudate nucleus and other non face-selective regions. Similar patterns emerge when comparing the saliency maps of the CNNs trained for 10 epochs, as only the saliency maps from the OS data set show elevated saliency merely for face-selective regions (compare section A.2) .

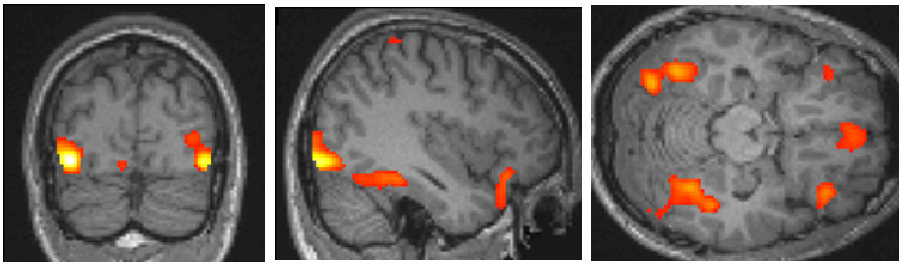
[3]The herein presented saliency map visualizations were created according to section A.1.



(a) Coronal slice (b) Sagittal slice (c) Axial slice

Figure 4.6: Saliency map from the NOS data set for participant 15 face condition, coronal, sagittal and axial slice at 40x-82x-16mm.

When comparing the saliency map (Figure 4.5) and the results of the classical statistical analysis via a general linear model (Figure 4.7) contrasting scrambled and face condition of participant 15, a high similarity is striking. Both point to the significance of FFA and especially OFA regions, the STS is not distinguishable in either.



(a) Coronal slice (b) Sagittal slice (c) Axial slice

Figure 4.7: Results of classical GLM of participant 15 at [40,-82,-16], overlayed on 15s MRI, corrected for family wise error at $p < 0.05$., at 40x-82x-16mm.

5 Discussion

In the present study, the possibility to use fast event-related fMRI data as input to a convolutional neural network was investigated. Data from a face paradigm was processed to create a data set with data-label pairs. The study aimed at constructing a processing pipeline and in doing so answer questions related to the data labelling as well as to the CNNs performance on the processed data. The data used comes from a multi-modal data set by Henson et al. (2003) and the CNN utilized for testing the processing is a task-based network, which has already been proven successful in single brain volume classification (Vu et al., 2020).

5.1 Creating data-label pairs from (task-based) fast event-related fMRI data

With the help of the shifting algorithm it was possible to create data-label pairs from fast event-related data by changing the trial definition. The brute force optimization further attempted to optimize this labelling by finding shifts which optimize the NSMD between scrambled and face scans.

The brute force method found the optimal time shifts to be between 3.7 and 6.4 seconds. These time shifts do **not** indicate the peak of the BOLD signal, they rather describe how much the trial onset must be shifted forward in order to optimize the alignment of the time components of fast event-related data. For the processed data this means that a scan which mainly captures the BOLD response to a face stimulus also receives the label face. Further, the captured response should be the highest possible, while not being overshadowed by the response to another stimulus. In the best case this is the peak of the hemodynamic response to the stimulus. For example, if a trial onset is shifted by circa 4 seconds, the scans included would have been collected 4 - 6 and possibly 6 – 8 seconds after the stimulus onset. If one refers back to Figure 3.1 a shift of 4 seconds would lead to volume three being labelled with the face condition, this matches the peak of the BOLD signal from the face stimulus.

On the one hand, the optimal shifts found in this study would lead to the corresponding condition inclusion of the BOLD signal peak, according to the definition of Logothetis and Wandell (2004). On the other hand, the peak would only sometimes be included when regarding the peaking times of 3-5 seconds set by Amaro and Barker (2006) and Hillman (2014). In case of the first trial, which had its onset at 0 seconds, the peak would be expected at 6-9 seconds when following the definition from Logothetis and Wandell (2004). A shift of around 5 seconds would lead to the scan from 6-8 seconds being labelled with the condition of the first trial, it would further have been collected in the peaking time of its hemodynamic response. If the

peak is, however, at 3-5 seconds after stimulus onset, as proposed by Amaro and Barker (2006) and Hillman (2014) a 5 second shift would lead to the exclusion of the scan capturing the peak signal. Investigating the peak timing in the FFA in particular, the time series analysis by Druzgal and D'Esposito (2003) revealed a BOLD signal peak in the FFA at roughly 6 seconds. A scan happening at 6 seconds being labelled with the corresponding stimulus condition is feasible for all shifts between, but not including, 2 and 6 seconds[1]. As all but one participant have an optimal shift in this time range, the results of this study can be considered in line with the results of Druzgal and D'Esposito (2003). The outlier is a 6.4 second shift, which would make it impossible to capture a 6 second peak. However, this shift might be greater due to personal difference in hemodynamic response (Logothetis & Wandell, 2004), as a 6.4 second shift would make a peak below 9 seconds possible and is therefore in line with description of the BOLD response by Logothetis and Wandell (2004). It is alternatively possible that the algorithm was not able to find the true optimal shift and only found a worse local optimum. This indicates that it cannot be guaranteed that the pipeline presented finds a global optimum. This is further evidenced by visual inspection of the complex non-smooth shape of the voxels time shift to NSMD curves like in Figure 4.1, as well as their mean, median and variance. All of the curves display a high amount of local minima and maxima. As further only 101 time shifts were tested and infinitely more time shifts exist in-between the set range of 0 to 10 seconds, it is to be expected that better optima exist. It is, nonetheless, infeasible to search for such an optimum as there is no guarantee that any optimum is the global optimum. Therefore, only a local optimum to the time shift can be found by the brute force method with the application of a variance measure, independent of the number of tried time shifts. Whether or not this also leads to the lack of an optimum in trial labelling is not certain as multiple time shifts could in theory lead to an optimal trial labelling. This is due to the fact that the trial labelling is, in our case, binary and not a smooth function. As long as in the end a correct label is given it is irrelevant if this was due to an optimal or less optimal shift. Further, depending on scale, it is possible that there is not just one optimal shift but rather multiple ones as one would not expect a difference between a 5.2 second shift and a 5.20001 second shift. The problem with a non smooth functions is the following: While small changes for the vast majority of cases do not influence the results, one particular tiny change can lead to a large difference. For example, if trial onset is 0 it would not matter which shift between 2 and 3.99999 is used, in all cases the scan happening between 4 and 6 seconds is labelled with the trials condition. If, however, the shift is 4.00001 seconds, the scan would be excluded, given that the scan started before the new shifted trial onset[2].

Summarizing, it was possible to create data-label pairs from fast event-related task-based fMRI data with the proposed pipeline. These data-labels pairs match the data-label pairs of Vu et al. (2020). Further, the shift used for optimize the labelling

[1] Disregarding the inclusion of the fixation time into a trials individual shift

[2] The current algorithm would also exclude the 4.0 second scan. However, it is debatable that an improved version of the algorithm should include it. Due to the fact that the trials had a fixation period which is included in the shift and further start within a scan, it is unlikely that a shifted trial onset is equal to a scan start, making this difference de facto irrelevant.

is a local optimum, however not a global one. If the resulting labelling is optimal can not be accessed as the underlying ground truth remains unknown.

5.2 Classification performance of a CNN for a (task-based) fast event-related fMRI input

The neural network classification performance on the test data set was very high and similar for both the data set with the optimal shift (99.37 %), as well as for the data set with a shift of 0 (100 %) when both networks performance were trained for 50 epochs. There is, however, a clear difference when the networks are trained for 10 epochs. The 91.80 % test accuracy for the optimal shift not only indicates that training for another 40 epochs might not be necessary, but further reveals a difference to the not optimal shift data set, which only had an accuracy of 77.85 %. This difference points to the OS data set facilitating faster training. Generally speaking, the classification accuracy is higher than that of Vu et al. (2020). This could be due to the lower number of categories as well as the higher amount of training data or due to the lack of a ground truth.

One reason leading to the high testing accuracy of the non-optimized data set might be the strength of the face effect. It is possible that the small variance between conditions at 0 – 4 seconds is enough for classification, if training is extensive. This could be explained by the particular shape of the BOLD signal, a possible initial dip, as well as any slope would still carry information distinguishing the conditions.

More importantly, the high performance of the NOS data set reveals a fundamental problem of the proposed pipeline when creating data sets for machine learning applications. The problem is the lack of a ground truth as the testing data is as correct or false as the training data. This means it is unknown whether a scan in fact shows the response to a certain condition and thereby is labelled correctly. Therefore, the true test accuracy of the artificial neural network remains unknown. Further, a wrong "true label" could lead to a false representation of the conditions within the network, while training. In this case, a neural network would learn a mapping of labels to data that does not reflect the true underlying biological difference of the conditions. This wrong mapping could make clinical applications which rely on biomarkers, such as diagnosis, impossible as the interpretability of the ANN is severely hampered (Yin et al., 2020).

Despite the lack of reliability of the reported test accuracy for judging the network's performance, inferences can be made about the CNNs performance by considering their saliency maps. The saliency map of the face condition for the optimal shift data shows high saliency in face-related regions such as FFA and OFA. This means the CNN uses activity in face-related areas for decision making. This is evidence for the CNN representing the categories face and scrambled accurately. The fact that higher OFA salience corresponds to stronger effects in the OFA in the classical GLM model results of participant 15 is further evidence for the CNN having correctly learned how the brain's response to a face stimulus looks like. For both the saliency map as well as the GLM no particular highlighting of the STS regions is evident. This indicates

that there is no relevant difference between the scrambled and face condition in participants 15s STS and, further, that the labelling through the optimal shift leads to the CNN learning the target concept of face processing correctly.

The face condition's saliency map for the zero shift data set fails to show the same face selectivity. Instead, there are points of heightened saliency scattered across the brain. Therefore, it cannot be assumed that the CNN uses face-related biomarker for decision making and learned the target concept of face viewing correctly. This further hints at the zero shift labelling leading to scans not capturing the key aspects of their label. Therefore, scans labelled face do not represent essential characteristics of the response to a face stimulus.

Vu et al. (2020) claim their CNN is able to extract task-relevant information from 3D fMRI volumes. On the one hand, saliency maps of the optimal shift data set support this notion, further showing its applicability to face paradigms. On the other hand, this feature extraction was not visible in the saliency map of the non optimal shift data set. Thereby, it can be assumed, that the CNN of Vu et al. (2020) is only able to extract task-relevant information from fMRI volumes if the labelling of those is adequate (enough).

5.3 The pipeline as a whole

As a whole was the pipeline successful in creating data-label pairs, which enable CNN training. The individual scripts can be used individually. For example training with data which went through different preprocessing is possible. This makes it possible to improve and adapt individual parts, such as the optimization, without disrupting the whole pipeline. The pipeline has an long execution time which roughly estimated lies over 40 hours without parallel computing, the optimization via brute force being the most time expensive unit.

5.4 Limitations of the present study and implications for the future

This study only investigated the specific fMRI data from the data set by Henson et al. (2003). This data set contains 16 young and healthy adult participant, and used a simple task paradigm. Therefore results might not translate to smaller data sets, different participant populations or other experimental paradigms. Further, only the specific CNN of Vu et al. (2020) was used for classification. Thus, results regarding classification might not translate to other artificial neural network (ANN)s, as well as other machine learning techniques using data-label pairs as input. Further, only binary classification was attempted and results might differ for classification tasks with more categories The CNN was only tested with one test subject, instead of using all participants as the test subject once, this limits conclusions regarding the generality of the pipeline.

Also, the exact relations of time shift to negative squared mean difference of the

two conditions still remain unclear. Further, conclusions drawn from and about the NSMD are influenced by the exclusion criteria, the underlying original trial definition, the number of trials and the fact that the exclusion of a scan is binary. Therefore, further investigation into which of these influences have which effects is necessary to make adequate correction possible or to at least acknowledge their effects in the future to get more accurate estimates of the optimal shift. Further, alternative optimization procedures could be investigated to not only improve the optimization results but also the execution time, this could for example be simulated annealing with adapted optimization parameters.

Moreover, the exclusion criteria should be inspected on their own as currently the data loss is at 29.25% in case of participant 15, meaning almost a third of all scrambled, famous and unfamiliar scans were excluded. The current exclusion criteria depend on the original three conditions famous, unfamiliar and scrambled. Later only the difference between face and scrambled condition is investigated, leading to the unnecessary exclusion of all scans which could not clearly be labelled as either famous or unfamiliar, but would nonetheless confidently be labelled face. Further, investigating the influence of exclusion criteria on the performance of the CNN would make it possible to use existing data more efficiently in the future.

The results of the CNN need to be regarded with caution. It is uncertain if the network used overfitts[3] as no analysis in that regard was done and learning curves were not monitored during training. Monitoring training would lead to more insights and better evaluation of the neural network's performance. To prevent overfitting in the future regularization, for example via an early stop criteria, could be employed, possibly leading to fewer epochs and shorter training times (cf. Goodfellow et al., 2016, Chapter 7).

In this study, the CNN was neither adapted nor optimized. The CNN implementation used, however, was not optimized for very large data sets and, thereby, can lead to memory problems. It was further implemented in the deprecated first version of TensorFlow. Updating this implementation would improve future applications of this network to large data sets. Further, no hyperparameter [4] optimization was done. Improving the hyperparameters for each application would be expensive but could improve classification results as the CNN learning schedule is tailored more closely to the learned problem.

The problem related to the missing ground truth is the most important one to address. It could be solved by creating an alternative testing set that is used with both the non and the optimized training data set. Such a data set could be built manually by carefully investigating the existing trial definition, looking for blocks of a condition, and selecting all scans which are certain to show the response to that condition. An alternative is to use data from both, a fast event-related design and a block design. An example for such a data can be found in the gambling task of the Human Connectome Project S1200 data set as both an event-related and a block design version is available (Barch et al., 2013; van Essen et al., 2012).

[3]Refer to Goodfellow et al., 2016, Chapter 5.2

[4]Parameters which control the learning (Goodfellow et al., 2016, Chapter 5.3).

The CNN could then be trained with the event-related data and tested with the block design data. Another alternative to fix the ground truth problem would be usage of synthetic fMRI data. With such data the ground truth is built through simulation and, therefore, known. Such a test against a ground truth, would ideally only needed to be done as proof of concept. A correct labelling resulting from it could be considered evidence for the validity of the suggested pipeline. That further could validate a pipeline's application to other data with missing ground truth.

Vu et al. (2020) were able to use their CNN for online classification. It must be tested if this is possible with fast event-related data. For such an application it would be necessary to find a sufficient shift that can be applied to the data in online classification a priory. Such a shift could be based on the optimal shifts of the training set.

Further, general applicability of this method has to be critically considered, as, despite the promising results, the CNN still suffers from the lack of insight into it. While saliency maps and other feature extraction methods can lead to insights about the workings of an ANN, the true meaning and the influence and meaning of the vast amount of parameters are still unclear. Therefore, it should be closely investigated for which application such a CNN is feasible and where other machine learning methods, with better known parameters, might perform better.

One possible application of this studies CNN trained on fast event-related face data could lie in aiding diagnosis of conditions which have shown to have a significant difference in face processing such as ASD or psychiatric disorders like bipolar disorder. This is possible, as once the CNN training is completed, classification of new data can be done in a matter of seconds. Moreover, the application of the proposed pipeline to alternative data and machine learning methods is feasible. In detail, experiments with paradigms, which can only be realized in a fast event-related paradigm can be used with machine learning methods, such as CNNs, which rely on data-label pairs. Using machine learning methods to investigate fMRI data could lead to discovering new aspects of information processing in the brain which could not be accessed before. Especially in the case of non-linear patterns, ANNs could provide new detection possibilities.

6 Conclusion

This study investigated the potential use of fast event-related fMRI data with artificial neural networks. For that, a processing pipeline was proposed and tested with fast event-related fMRI of a face paradigm by Henson et al. (2003) and a CNN proposed by Vu et al. (2020). This pipeline consists of the steps preprocessing, optimization, creating a new trial definition, normalization and data set creation. This pipeline was successful in creating a data set with data-label pairs by aligning the time components of fast event-related data and excluding scans to which no definite label could be assigned. Only the CNN trained on the optimized data-label pairs successfully revealed task relevant brain areas, evidencing the necessity of such an optimization step. However, more evidence is needed to ensure that the labelling created by the pipeline is correct. This includes a test against a ground truth to reveal the true test performance of the trained CNN.

Overall, this study provides evidence for the adequacy of using fast event-related fMRI data as input to machine learning applications requiring data-label pairs. However, it also showed that more research is needed before an artificial neural network trained on fast event-related data-label pairs can be used with more complex paradigms, as a stand-alone analysis tool or to aid clinical diagnosis.

Bibliography

Amaro, E., & Barker, G. J. (2006). Study design in fmri: Basic principles. *Brain and cognition*, *60*(3), 220–232. https://doi.org/10.1016/j.bandc.2005.11.009

Barch, D. M., Burgess, G. C., Harms, M. P., Petersen, S. E., Schlaggar, B. L., Corbetta, M., Glasser, M. F., Curtiss, S., Dixit, S., Feldt, C., Nolan, D., Bryant, E., Hartley, T., Footer, O., Bjork, J. M., Poldrack, R., Smith, S., Johansen-Berg, H., Snyder, A. Z., & van Essen, D. C. (2013). Function in the human connectome: Task-fmri and individual differences in behavior. *NeuroImage*, *80*, 169–189. https://doi.org/10.1016/j.neuroimage.2013.05.033

Bennett, C. M., Miller, M. B., & Wolford, G. L. (2009). Neural correlates of interspecies perspective taking in the post-mortem atlantic salmon: An argument for multiple comparisons correction. *NeuroImage*, *47*, S125. https://doi.org/10.1016/S1053-8119(09)71202-9

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.

Burock, M. A., Buckner, R. L., Woldorff, M. G., Rosen, B. R., & Dale, A. M. (1998). Randomized event-related experimental designs allow for extremely rapid presentation rates using functional mri. *Neuroreport*, *9*(16), 3735–3739. https://doi.org/10.1097/00001756-199811160-00030

Collins, J. A., & Olson, I. R. (2014). Beyond the ffa: The role of the ventral anterior temporal lobes in face processing. *Neuropsychologia*, *61*, 65–79. https://doi.org/10.1016/j.neuropsychologia.2014.06.005

Colman, A. M. (2009). *A dictionary of psychology* (3rd ed.). Oxford University Press. https://doi.org/10.1093/acref/9780199534067.001.0001

Davies-Thompson, J., Newling, K., & Andrews, T. J. (2013). Image-invariant responses in face-selective regions do not explain the perceptual advantage for familiar face recognition. *Cerebral cortex (New York, N.Y. : 1991)*, *23*(2), 370–377. https://doi.org/10.1093/cercor/bhs024

Di Visconti Oleggio Castello, M., Halchenko, Y. O., Guntupalli, J. S., Gors, J. D., & Gobbini, M. I. (2017). The neural representation of personally familiar and unfamiliar faces in the distributed system for face perception. *Scientific reports*, *7*(1), 12237. https://doi.org/10.1038/s41598-017-12559-1

Drew, P. J. (2019). Vascular and neural basis of the bold signal. *Current Opinion in Neurobiology*, *58*, 61–69. https://doi.org/10.1016/j.conb.2019.06.004

Druzgal, T. J., & D'Esposito, M. (2003). Dissecting contributions of prefrontal cortex and fusiform face area to face working memory. *Journal of Cognitive Neuroscience*, *15*(6), 771–784. https://doi.org/10.1162/089892903322370708

Ebrahimighahnavieh, A., Luo, S., & Chiong, R. (2020). Deep learning to detect alzheimer's disease from neuroimaging: A systematic literature review. *Computer Methods and Programs in Biomedicine*, *187*, 105242. https://doi.org/10.1016/j.cmpb.2019.105242

Glover, G. H. (2011). Overview of functional magnetic resonance imaging. *Neurosurgery clinics of North America*, *22*(2), 133–9, vii. https://doi.org/10.1016/j.nec.2010.11.001

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning* [http://www.deeplearningbook.org]. MIT Press.

Henson, R. N., Goshen-Gottstein, Y., Ganel, T., Otten, L. J., Quayle, A., & Rugg, M. D. (2003). Electrophysiological and haemodynamic correlates of face perception, recognition and priming. *Cerebral cortex (New York, N.Y. : 1991)*, *13*(7), 793–805. https://doi.org/10.1093/cercor/13.7.793

Hillman, E. M. C. (2014). Coupling mechanism and significance of the bold signal: A status report. *Annual review of neuroscience*, *37*, 161–181. https://doi.org/10.1146/annurev-neuro-071013-014111

Huettel, S. A. (2012). Event-related fmri in cognition. *NeuroImage*, *62*(2), 1152–1156. https://doi.org/10.1016/j.neuroimage.2011.08.113

Huettel, S. A., Song, A. W., & McCarthy, G. (2008). *Functional magnetic resonance imaging* (2nd ed.). W. H. Freeman; Basingstoke : Palgrave [distributor].

Kim, S.-G., & Bandettini, P. A. (2012). Principles of bold functional mri. In S. H. Faro, F. B. Mohamed, M. Law, & J. T. Ulmer (Eds.), *Functional neuroradiology* (pp. 293–303). Springer US. https://doi.org/10.1007/978-1-4419-0345-7{\textunderscore}16

Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278–2324. https://doi.org/10.1109/5.726791

Lee, D., Yun, S., Jang, C., & Park, H.-J. (2017). Multivariate bayesian decoding of single-trial event-related fmri responses for memory retrieval of voluntary actions. *PloS one*, *12*(8), e0182657. https://doi.org/10.1371/journal.pone.0182657

Logothetis, N. K., & Wandell, B. A. (2004). Interpreting the bold signal. *Annual Review of Physiology*, *66*(1), 735–769. https://doi.org/10.1146/annurev.physiol.66.082602.092845

Mumford, J. A., Turner, B. O., Ashby, F. G., & Poldrack, R. A. (2012). Deconvolving bold activation in event-related designs for multivoxel pattern classification analyses. *NeuroImage*, *59*(3), 2636–2643. https://doi.org/10.1016/j.neuroimage.2011.08.076

Ogawa, S., Lee, T. M., Kay, A. R., & Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences of the United States of America*, *87*(24), 9868–9872. https://doi.org/10.1073/pnas.87.24.9868

Rhodes, G., Calder, A., Johnson, M., & Haxby, J. V. (2012). *Oxford handbook of face perception*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199559053.001.0001

Riaz, A., Asad, M., Alonso, E., & Slabaugh, G. (2020). Deepfmri: End-to-end deep learning for functional connectivity and classification of adhd using fmri. *Journal of neuroscience methods*, *335*, 108506. https://doi.org/10.1016/j.jneumeth.2019.108506

Rorden, C., & Brett, M. (2000). Stereotaxic display of brain lesions. *Behavioural neurology*, *12*(4), 191–200. https://doi.org/10.1155/2000/421719

Rosen, B. R., Buckner, R. L., & Dale, A. M. (1998). Event-related functional mri: Past, present, and future. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(3), 773–780. https://doi.org/10.1073/pnas.95.3.773

Schirrmeister, R. T., Springenberg, J. T., Fiederer, L. D. J., Glasstetter, M., Eggensperger, K., Tangermann, M., Hutter, F., Burgard, W., & Ball, T. (2017). Deep learning with convolutional neural networks for eeg decoding and visualization. *Human brain mapping*, *38*(11), 5391–5420. https://doi.org/10.1002/hbm.23730

Skrandies, W. (1990). Global field power and topographic similarity. *Brain topography*, *3*(1), 137–141. https://doi.org/10.1007/BF01128870

The SPM Developers. (2020). Spm12 manual. https://www.fil.ion.ucl.ac.uk/spm/doc/manual.pdf

van Essen, D. C., Ugurbil, K., Auerbach, E., Barch, D., Behrens, T. E. J., Bucholz, R., Chang, A., Chen, L., Corbetta, M., Curtiss, S. W., Della Penna, S., Feinberg, D., Glasser, M. F., Harel, N., Heath, A. C., Larson-Prior, L., Marcus, D., Michalareas, G., Moeller, S., ... Yacoub, E. (2012). The human connectome project: A data acquisition perspective. *NeuroImage*, *62*(4), 2222–2231. https://doi.org/10.1016/j.neuroimage.2012.02.018

Vu, H., Kim, H.-C., Jung, M., & Lee, J.-H. (2020). Fmri volume classification using a 3d convolutional neural network robust to shifted and scaled neuronal activations. *NeuroImage*, 117328. https://doi.org/10.1016/j.neuroimage.2020.117328

Yarkoni, T., Poldrack, R. A., Nichols, T. E., van Essen, D. C., & Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nature methods*, *8*(8), 665–670. https://doi.org/10.1038/nmeth.1635

Yin, W., Li, L., & Wu, F.-X. (2020). Deep learning for brain disorder diagnosis based on fmri images. *Neurocomputing*. https://doi.org/10.1016/j.neucom.2020.05.113

Zhang, C., Qiao, K., Wang, L., Li Tong, Hu, G., Zhang, R.-Y., & Yan, B. (2019). A visual encoding model based on deep neural networks and transfer learning for brain activity measured by functional magnetic resonance imaging. *Journal of neuroscience methods*, *325*, 108318. https://doi.org/10.1016/j.jneumeth.2019.108318

Zhao, X., & Zhao, X.-M. (2020). Deep learning of brain magnetic resonance images: A brief review. *Methods (San Diego, Calif.)* https://doi.org/10.1016/j.ymeth.2020.09.007

Zubarev, I., Zetter, R., Halme, H.-L., & Parkkonen, L. (2019). Adaptive neural network classifier for decoding meg signals. *NeuroImage*, *197*, 425–434.

A Saliency maps

A.1 Creation of the saliency map visualization

The saliency maps obtained from the CNN were saved into a NIfTI file. Followingly, they were overlayed on participant 15s structural to MNI15 space aligned MRI (named "wmprage") by using MRIcro and an "ACTC" colormap with 40% transparency (Rorden & Brett, 2000).

A.2 Other saliency maps

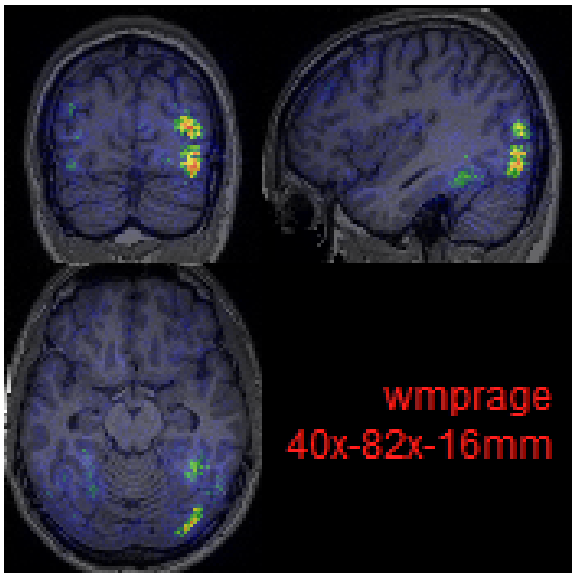


Figure A.1: Saliency map from the OS data set for participant 15 scrambled face condition, coronal, sagittal and axial slice at 40x-82x-16mm.

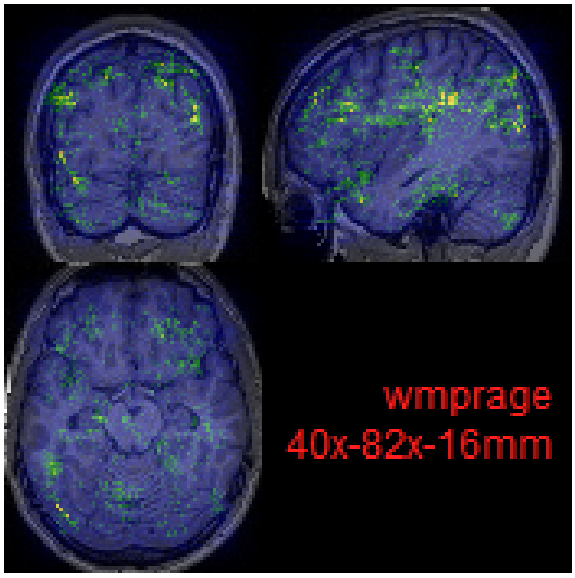


Figure A.2: Saliency map from the NOS data set for participant 15 scrambled face condition, coronal, sagittal and axial slice at 40x-82x-16mm.

A.3 Saliency maps for 10 epochs

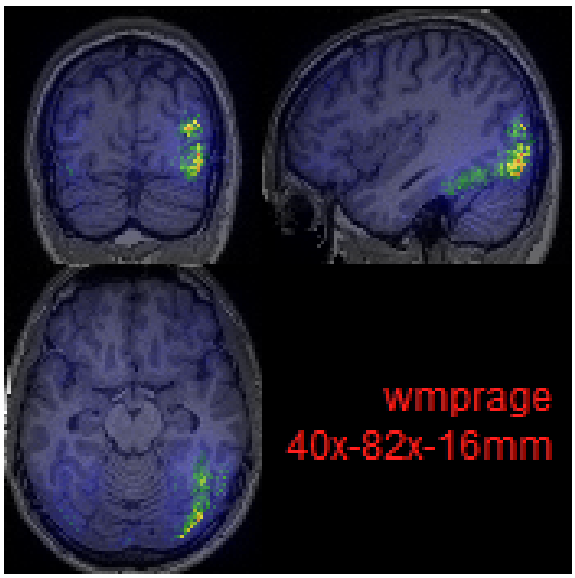


Figure A.3: Saliency map from the OS data set for participant 15 face condition after 10 epochs, coronal, sagittal and axial slice at 40x-82x-16mm.

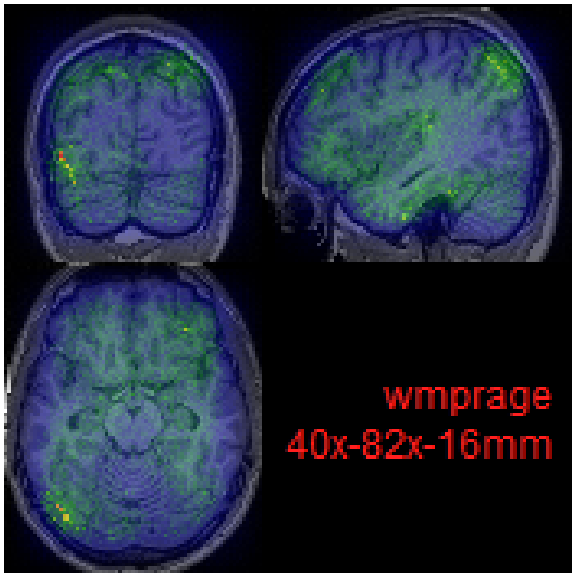


Figure A.4: Saliency map from the NOS data set for participant 15 face condition after 10 epochs, coronal, sagittal and axial slice at 40x-82x-16mm.

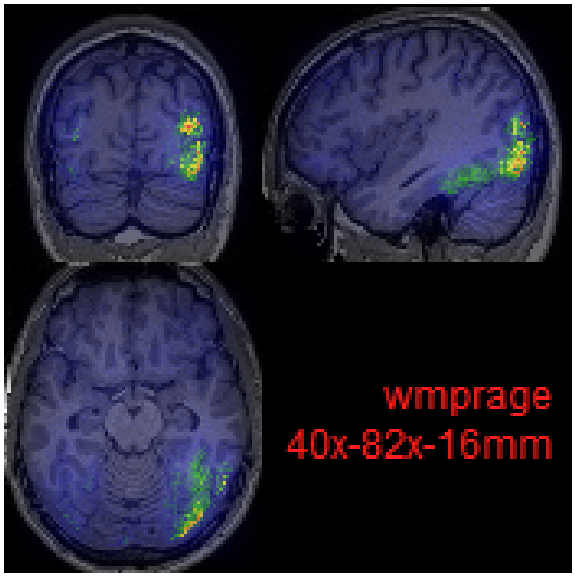


Figure A.5: Saliency map from the OS data set for participant 15 scrambled condition after 10 epochs, coronal, sagittal and axial slice at 40x-82x-16mm.

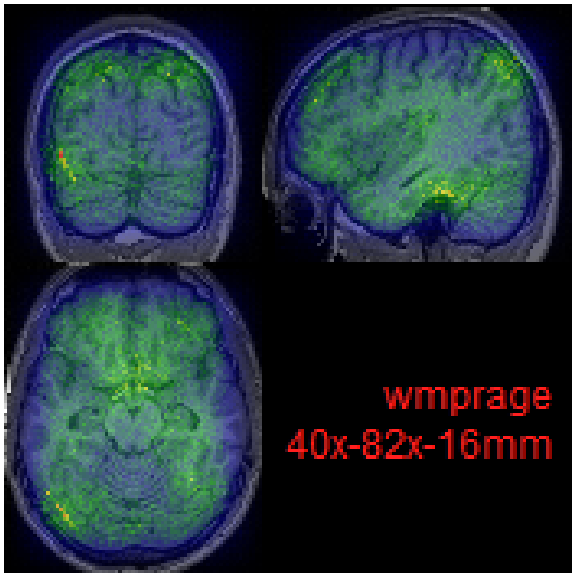


Figure A.6: Saliency map from the NOS data set for participant 15 scrambled condition after 10 epochs, coronal, sagittal and axial slice at 40x-82x-16mm.

B Extra Material

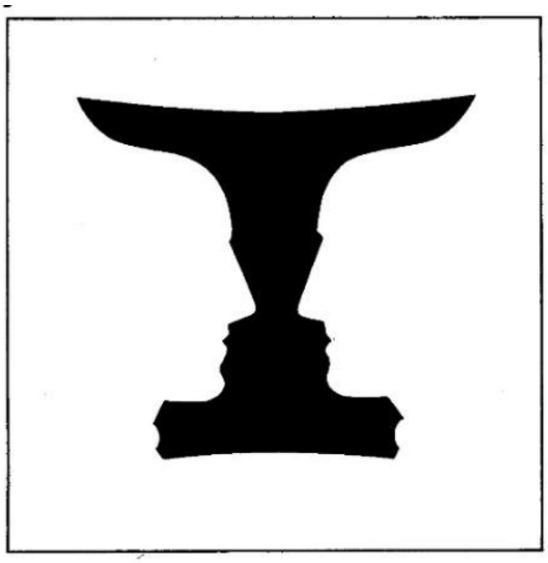


Figure B.1: An example of a Rubin's vase image. Retrieved from: Using a Conscious System to Construct a Model of the Rubin's Vase Phenomenon - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/An-example-of-a-Rubins-Vase-image_fig1_309182998.

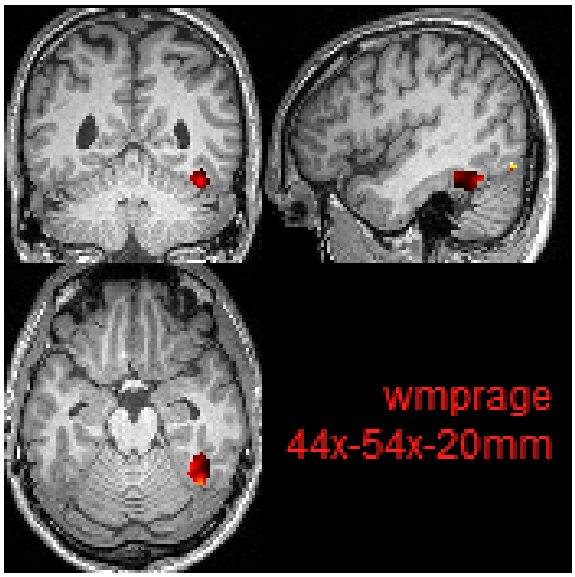


Figure B.2: Range in ROI voxels in participant 15, high range in OFA.

Declaration of Authorship

I hereby certify that the work presented here is, to the best of my knowledge and belief, original and the result of my own investigations, except as acknowledged, and has not been submitted, either in part or whole, for a degree at this or any other university.

_____

Signature

_____

City, Date