CARDIFF UNIVERSITY

DOCTORAL THESIS

From features to concepts: tracking the neural dynamics of visual perception

Diana C. DIMA

Supervisor:
Prof. Krish D. SINGH

A thesis submitted in fulfillment of the requirements
for the degree of Doctor of Philosophy

October 2018

Abstract

From features to concepts: tracking the neural dynamics of visual perception

The visual system is thought to accomplish categorization through a series of hierarchical feature extraction steps, ending with the formation of high-level category representations in occipitotemporal cortex; however, recent evidence has challenged these assumptions. The experiments described in this thesis address the question of categorization in face and scene perception using magnetoencephalography and multivariate analysis methods.

The first three chapters investigate neural responses to emotional faces from different perspectives, by varying their relevance to task. First, in a passive viewing paradigm, angry faces elicit differential patterns within 100 ms in visual cortex, consistent with a threat-related bias in feedforward processing. The next chapter looks at rapid face perception in the context of an expression discrimination task which also manipulates subjective awareness. A neural response to faces, but not expressions is detected outside awareness. Furthermore, neural patterns and behavioural responses are shown to reflect both facial features and facial configuration. Finally, the third chapter employs emotional faces as distractors during an orientation discrimination task, but finds no evidence of expression processing outside of attention.

The fourth chapter focuses on natural scene perception, using a passive viewing paradigm to study the contribution of low-level features and high-level categories to MEG patterns. Multivariate analyses reveal a categorical response to scenes emerging within 200 ms, despite ongoing processing of low-level features.

Together, these results suggest that feature-based coding of categories, optimized for both stimulus relevance and task demands, underpins dynamic high-level representations in the visual system. The findings highlight new avenues in vision research, which may be best pursued by bridging the neural and behavioural levels within a common computational framework.

Declaration and Statements

Declaration

This work has not been submitted in substance for any other degree or award at this or any other university or place of learning, nor is being submitted concurrently in candidature for any degree or other award.

Signed: _____ (candidate) Date: _____

Statement 1

This thesis is being submitted in partial fulfillment of the requirements for the degree of PhD.

Signed: _____ (candidate) Date: _____

Statement 2

This thesis is the result of my own independent work/investigation, except where otherwise stated, and the thesis has not been edited by a third party beyond what is permitted by Cardiff University's Policy on the Use of Third Party Editors by Research Degree Students. Other sources are acknowledged by explicit references. The views expressed are my own.

Signed: _____ (candidate) Date: _____

Statement 3

I hereby give consent for my thesis, if accepted, to be available online in the University's Open Access repository and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed: _____ (candidate) Date: _____

Acknowledgements

There are a number of people without whom this thesis would not have been possible. I would first like to thank my supervisor, Krish Singh, for being incredibly supportive throughout these three years. His guidance and encouragement allowed me to find and tackle the topics most exciting to me, and I've learned so much from his approach to research.

I would also like to thank Gavin Perry for his help in starting out with MEG and throughout my PhD, and Lorenzo Magazzini, who fielded endless questions about MEG analysis and navigated version control with me. Thanks also go to Jiaxiang Zhang for his help with MVPA analysis, and to Eirini Messaritaki, who collected the first experimental dataset in this thesis together with me. Thanks to everyone in the MEG lab for valuable feedback and discussions. A special shout-out to Sophie Esterer, who helped me navigate PhD life and thesis writing, and reminded me to get outdoors.

There are too many people to name in CUBRIC to whom I am thankful for help, interesting collaborations, or much-needed coffee breaks. But special thanks must go to Cyril Charron and the IT team for all their help and hard work setting up the CUBRIC systems. I would also like to thank everyone who took part in the MEG studies in this thesis, including the endless piloting sessions, with lots of patience and minimal head motion.

A huge thank you goes to my family and friends, near and far, with whom I've shared great trips and conversations, and who have made these years amazing. Thanks to my brother for making the trip to Cardiff and brightening some revision-filled days during my PhD.

And finally, to Akshay, for your support day in and day out, and sharing in all the ups and downs of these three years: you've made it all so much easier.

Impact of this thesis

The results of Chapters 2 and 5 have been published as peer-reviewed journal articles:

- Dima, D.C, Perry, G., Messaritaki, E., Zhang, J., & Singh, K. D. (2018). Spatiotemporal dynamics in human visual cortex rapidly encode the emotional content of faces. *Human Brain Mapping* 39(10):3993-4006. 10.1002/hbm.24226.

- Dima, D.C., Perry, G., & Singh, K.D. (2018). Spatial frequency supports the emergence of categorical representations in visual cortex during natural scene perception. *NeuroImage* 179:102–116. 10.1016/j.neuroimage.2018.06.033.

The results of Chapter 3 are available as a pre-print and have been submitted for publication:

- Dima, D.C. & Singh, K.D. (submitted). Dynamic representations of behaviourally relevant features support rapid face processing in the human ventral visual stream. *biorXiv pre-print*: 10.1101/394916.

Data collection

All analyses described in this thesis were performed by me.
The data presented in Chapter 2 were collected jointly by Eirini Messaritaki and me. All other datasets presented in this thesis were collected by me.

Contents

List of Figures

List of Tables

List of Abbreviations

AAL	**A**utomated **A**natomical **L**abeling
BOLD	**B**lood **O**xygen **L**evel **D**ependent
CNN	**C**onvolutional **N**eural **N**etwork
DNN	**D**eep **N**eural **N**etwork
EEG	Electro**e**ncephalo**g**raphy
EOG	Electro**o**culo**g**raphy
ERF	**E**vent-**R**elated **F**ield
ERP	**E**voked **R**esponse **P**otential
FFA	**F**usiform **F**ace **A**rea
fMRI	**F**unctional **M**agnetic **R**esonance **I**maging
LCMV	**L**inearly **C**onstrained **M**inimum **V**ariance
MDS	**M**ulti-**D**imensional **S**caling
MEG	**M**agneto**e**ncephalo**g**raphy
MRI	**M**agnetic **R**esonance **I**maging
MVPA	**M**ulti**v**ariate **P**attern **A**nalysis
OFA	**O**ccipital **F**ace **A**rea
OPA	**O**ccipital **P**lace **A**rea
PPA	**P**arahippocampal **P**lace **A**rea
RDM	**R**epresentational **D**issimilarity **M**atrix
ROI	**R**egion **o**f **I**nterest
RSA	**R**epresentational **S**imilarity **A**nalysis
RSC	**R**etro**s**plenial **C**ortex
SNR	**S**ignal-to-**N**oise **R**atio
STS	**S**uperior **T**emporal **S**ulcus
SVD	**S**ingular **V**alue **D**ecomposition
SVM	**S**upport **V**ector **M**achine

Chapter 1

General introduction

1.1 The puzzle of human vision

In daily life, we are hardly aware of the processes that lead to us perceiving, understanding, and acting upon what we perceive. Visual perception, though apparently effortless, has turned out to be difficult to unravel or implement in artificial systems, although significant progress has been made in recent years. The immensity of this task becomes apparent as soon as we consider the transformations involved: from a virtually infinite set of possible light signals reflected by any given stimulus across viewing conditions, to the accurate categorization of that stimulus (Cox, 2014).

This complex and variable visual information is captured by photoreceptor cells in the retina, which transmit it to the visual system via ganglion cells in the optic nerve. In the primary visual cortex (V1), these outputs are pooled by neurons with highly selective receptive fields, tuned to local edges of specific orientations (Hubel and Wiesel, 1962). This selectivity continues throughout the retinotopically organized extrastriate visual cortex (Bullier, 2001) with increasingly complex features, and turns into a broader category selectivity in the ventral temporal cortex. This is where a progression from "low-level" to "high-level" vision is commonly proposed: while neurons in early visual areas respond to local visual features, ventral stream areas are thought to encode a range of mid-level features or high-level categories, including colour, object category, object size, concepts, etc. (Grill-Spector and Weiner, 2014). The computations performed in occipitotemporal cortex at this later stage have been the subject of significant debate. Understanding whether

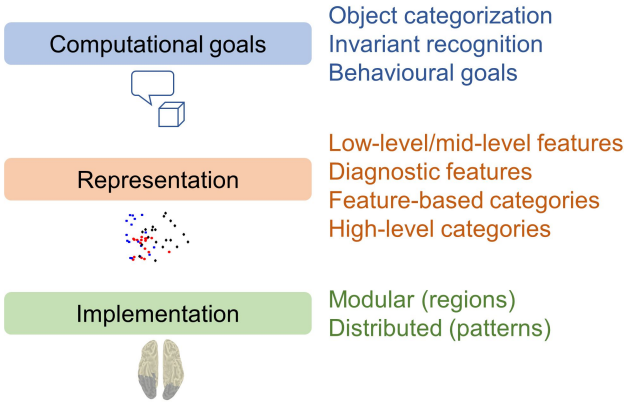

FIGURE 1.1: Framework for understanding high-level vision according to Marr's schema, together with some proposed solutions. The four types of representations are adapted from those suggested by Bracci et al. (2017).

ventral areas represent visual features or abstract categories would help answer broader questions about the role of modality-specific brain areas in the emergence of conceptual knowledge (Bracci et al., 2017a), and ultimately about the interface between sensory and semantic information, or between perception and cognition (Beck, 2018).

1.1.1 Goals in high-level vision

The functions of the ventral visual stream have sometimes been framed according to Marr's threefold schema for understanding information processing systems (Marr, 1982): its computational goals, the representations it employs, and their implementation or neural substrates (Fairhall, 2014; Grill-Spector and Weiner, 2014). However, little agreement has been reached on what the three elements might be in the high-level visual system (Figure 1.1).

One of the most influential principles in human vision is that of a dual-pathway architecture, consisting of two interacting ventral and dorsal visual streams. The two systems are thought to perform visual processing for perception and action respectively (Goodale and Milner, 1992), or to separately extract object identity and spatial information (Ungerleider and Haxby, 1994). Within this framework, hierarchical models of the ventral visual stream usually adopt an object recognition perspective, whereby the goal of the system is categorization, understood as the matching of stimulus representations to object representations stored in long-term

memory (Bracci et al., 2017a). This can be achieved through an efficient and explicit organization of category-specific neural representations, and through invariance to visual and cross-exemplar variability. In this sense, the hierarchy of visual processing has been described as an "untangling" of categories from the corresponding visual features through a series of linear and non-linear operations (Grill-Spector and Weiner, 2014; Rust and DiCarlo, 2010), with the stimulus being effectively "decoded" at the end of this process. This view has been reinforced in recent years by the success of feedforward neural networks in solving object recognition tasks, which showed that category selectivity can arise from a series of combinations of visual features (Jozwik et al., 2016; Peelen and Downing, 2017).

However, it has been argued that an object recognition framework is an oversimplification of what the visual system needs to accomplish, given the large variety of inputs and tasks we encounter daily (Cox, 2014; Peelen and Downing, 2017). Object categories themselves are complex, ranging from taxonomic to functional (Bracci et al., 2017a) and from specific to abstract (Edelman et al., 1998). Furthermore, such categories need to be adapted to specific behavioural goals (Groen et al., 2017), and their representations will necessarily vary due to differences in relevant features (Figure 1.2). In this sense, the processing of a scene in the ventral visual stream will depend on whether the goal is navigation (e.g. detection of affordances; Bonner and Epstein, 2017; Epstein, 2008), assessing social information (body cues; Downing and Peelen, 2011), or recognizing somebody based on their face (configural face processing; Freiwald et al., 2016). Rather than making a distinction between low-level features and high-level categories, it might be better to investigate behaviourally relevant features during naturalistic tasks (Peelen and Downing, 2017).

These perspectives highlight two different accounts of what the visual system is optimized to accomplish: a purely visual category selectivity, where stimuli are categorized in the visual system, but further assessment happens at later stages of cognition (Kravitz et al., 2013), and a conceptual selectivity, influenced by previous knowledge and task demands and not restricted to visual features (Bracci and Beeck, 2016; Kaiser et al., 2016). These accounts determine the types of neural representations thought to support these goals in occipitotemporal cortex.

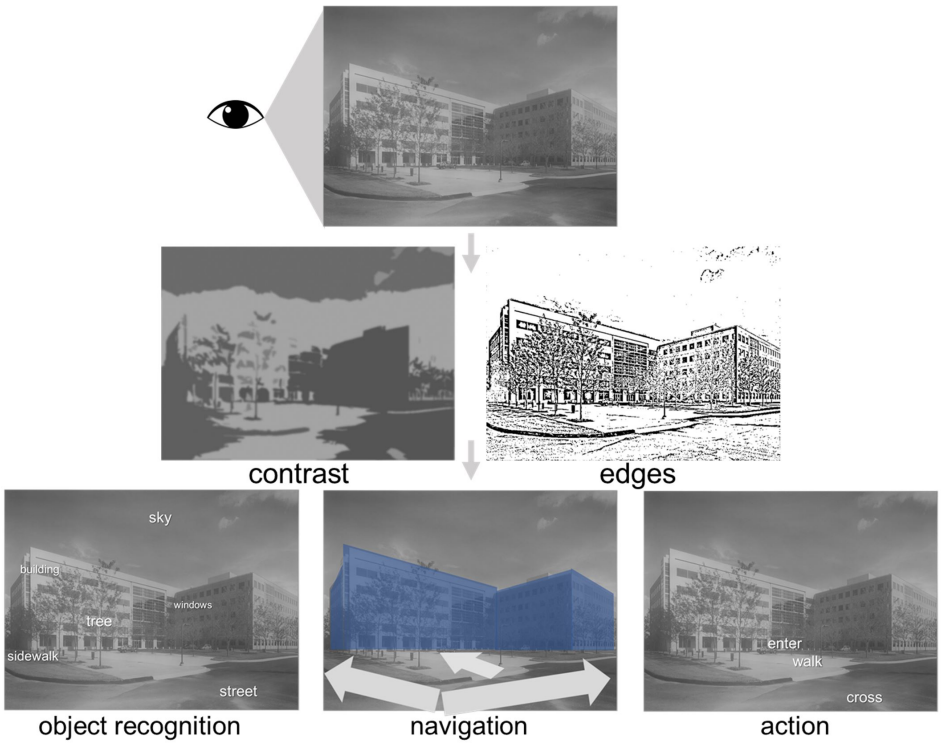

FIGURE 1.2: In the progression from low-level features to high-level representations, task and context can shape the type of features extracted and represented. Some examples of low-level features and high-level behaviourally relevant representations are shown.

1.1.2 Features and categories in ventral stream representations

The presence of category-selective responses in the visual ventral stream has been well-documented, starting with neuropsychological investigations (e.g. Sacchett and Humphreys, 1992; Warrington and Shallice, 1984) and continuing with a wealth of neuroimaging studies (e.g. Bell et al., 2009; Carlson et al., 2003; Haxby et al., 2001; Hung et al., 2005; Kriegeskorte et al., 2008). However, visual features often correlate with high-level categories and are not always controlled (Cox and Savoy, 2003). High-level visual areas have been shown to respond to low-level and mid-level visual features (Andrews et al., 2015; Baldassi et al., 2013; Beeck et al., 2008; Caldara et al., 2006; Haxby et al., 2000; Ishai et al., 1999; Long et al., 2018; Nasr and Tootell, 2012; Nasr et al., 2014; Rajimehr et al., 2011; Rice et al., 2014; Woodhead et al., 2011). Seemingly conflicting results showing both invariant category selectivity and visual feature processing in the ventral stream can be resolved by adopting a feature-based account of category coding (Bracci et al., 2017a). Evidence of overlapping visual and categorical representations (Hong et al., 2016; Ramkumar et al.,

2016; Chapter 5) points to the role played by diagnostic visual features in the formation of high-level representations.

A recent review (Bracci et al., 2017a) grouped hypotheses about the content of such representations into four categories: low-level feature coding (exclusively visual), abstract category coding (exclusively high-level), diagnostic featural coding (representations of features characteristic of categories), and feature-based categorical coding (entailing both feature and category effects). The wealth of evidence showing both visual and categorical representations in the ventral visual stream suggests that the latter two are the most plausible hypotheses. Furthermore, assessing the relationship between such representations and behavioural responses can help uncover whether the categories are task-relevant (Carlson et al., 2013; Ritchie and Carlson, 2016; Tong and Pratte, 2012), and whether behavioural goals influence representations in the visual system.

1.1.3 The timing of categorization

This brings us to a related question: if high-level vision is a highly adaptable process optimized to accomplish behavioural and categorization goals across a range of viewing conditions and visual properties, how early does this optimization start?

The debate on the boundaries of perception and cognition (or so-called cognitive penetrability; Newen and Vetter, 2017) can be reframed as a debate on top-down influences on perception both within and outside of the visual system (Teufel and Nanay, 2017). While classic models envisioned a feedforward information flow converting features into high-level representations, more recent evidence has highlighted an important role of feedback connections at all stages of vision (Bar et al., 2006; Bullier, 2001; Gilbert and Li, 2013; Lamme and Roelfsema, 2000). In contrast with the hierarchical view, feedback connections have been shown to modulate neuronal tuning and neuronal population dynamics according to object expectations, context and task-related changes (Gilbert and Li, 2013). Category knowledge and learning shapes visual feature representations at the earliest stages of vision (Folstein et al., 2014, 2015; Teufel, 2018).

What is more, the role of prior information in shaping perception is not limited to top-down influences: constraints based on evolutionarily-relevant or naturally

occurring stimuli can be placed on visual perception and affect the extraction of relevant features (Teufel and Nanay, 2017). Evidence of categorical and contextual effects on early vision ties in with a model of adaptable, feature-based category coding in the visual system, optimized to create sparse representations guided by stimulus relevance and current behavioural goals.

1.1.4 Neural substrates: towards information mapping

A computational model of high-level vision also requires an understanding of how its algorithms and representations are implemented within the constraints of brain structure (Grill-Spector and Weiner, 2014). Ideas about how the ventral stream encodes category selectivity have changed in time from a modular view of functionally specialized regions (e.g. Epstein and Kanwisher, 1998; Kanwisher et al., 1997) to an information-based account of distributed representations (Kriegeskorte et al., 2007, 2008). It is thought that the separability of category-specific representations is achieved at different spatial scales, through functional clustering of neurons within columns, patches, regions and maps, and through topological features that are consistent across subjects. Furthermore, overlapping representations of different categories point to information integration as a mechanism to increase efficiency (Grill-Spector and Weiner, 2014).

Given the high dimensionality of these representations (Haxby et al., 2011), they can be approached either through model-based simplifications, or through data-driven information mapping techniques that can uncover the underlying lower-dimensional structures (Bracci et al., 2017a; Fairhall, 2014; Sussillo, 2014). At a time when machine learning is ready to move from object recognition to natural behaviour (Fairhall, 2014), we may be able to uncover the axes separating high-level representations in the visual system by combining pattern recognition, rich neuroimaging data, and careful experimental design for maximal interpretability.

1.2 Recording neural activity with MEG

In this thesis, the question of high-level vision (face and scene perception) is addressed by combining experimental designs that manipulate visual properties and

behaviour with magnetoencephalography (MEG) and information mapping techniques. Over the past decade, the application of pattern recognition to neuroimaging has revolutionized the field (Haxby et al., 2001; Haxby et al., 2014; Kamitani and Tong, 2005). Applying these methods to electrophysiological recordings (MEG, EEG or intracranial EEG) has also become more and more common: the timing of neural processes can provide a window into the underlying mechanisms, and increasingly sophisticated multivariate techniques have been used to resolve their temporal dynamics. This thesis focuses on the use of MEG to capture the complex whole-brain dynamics of high-level perceptual processing, and on pattern recognition as a method of resolving them within a data-driven framework.

1.2.1 Neuronal generators and MEG instrumentation

MEG offers a non-invasive measure of the magnetic fields produced by electrical currents in the brain. Although neural electric activity comprises both rapid action potentials and slower synaptic potentials, intracellular post-synaptic potentials generated at the apical dendrites of pyramidal neurons are thought to make the largest contribution to MEG signals (Baillet et al., 2001; Hari and Salmelin, 2012; Silva, 2010; Figure 1.3A). To be detectable with MEG, the firing of thousands of spatially aligned neurons needs to synchronize, such that the superposition of neural currents produces a measurable magnetic field (Baillet et al., 2001). Cortical pyramidal neurons are organized in palisades and perpendicular to the cortical surface (Nunez and Silberstein, 2000), thus forming "open fields" (No, 1947) and behaving as effective current dipoles (Silva, 2010). Thus, slower potentials generated at their dendrites are more likely to contribute to the MEG signal than rapid action potentials, which are unlikely to synchronize on a sufficient scale and whose magnetic fields decay more rapidly with distance (Singh, 2006).

Magnetic fields generated by the brain are extremely weak (50-500 fT; Hämäläinen et al., 1993). Although the first human MEG recording was made with a conventional coil (Cohen, 1968), sensitive measurements of these weak fields require superconducting quantum interference devices (SQUIDs; Cohen, 1972; Zimmerman et al., 1970). These are small coils which become superconducting when immersed in liquid helium with a temperature of approximately -270°C (Singh, 2006).

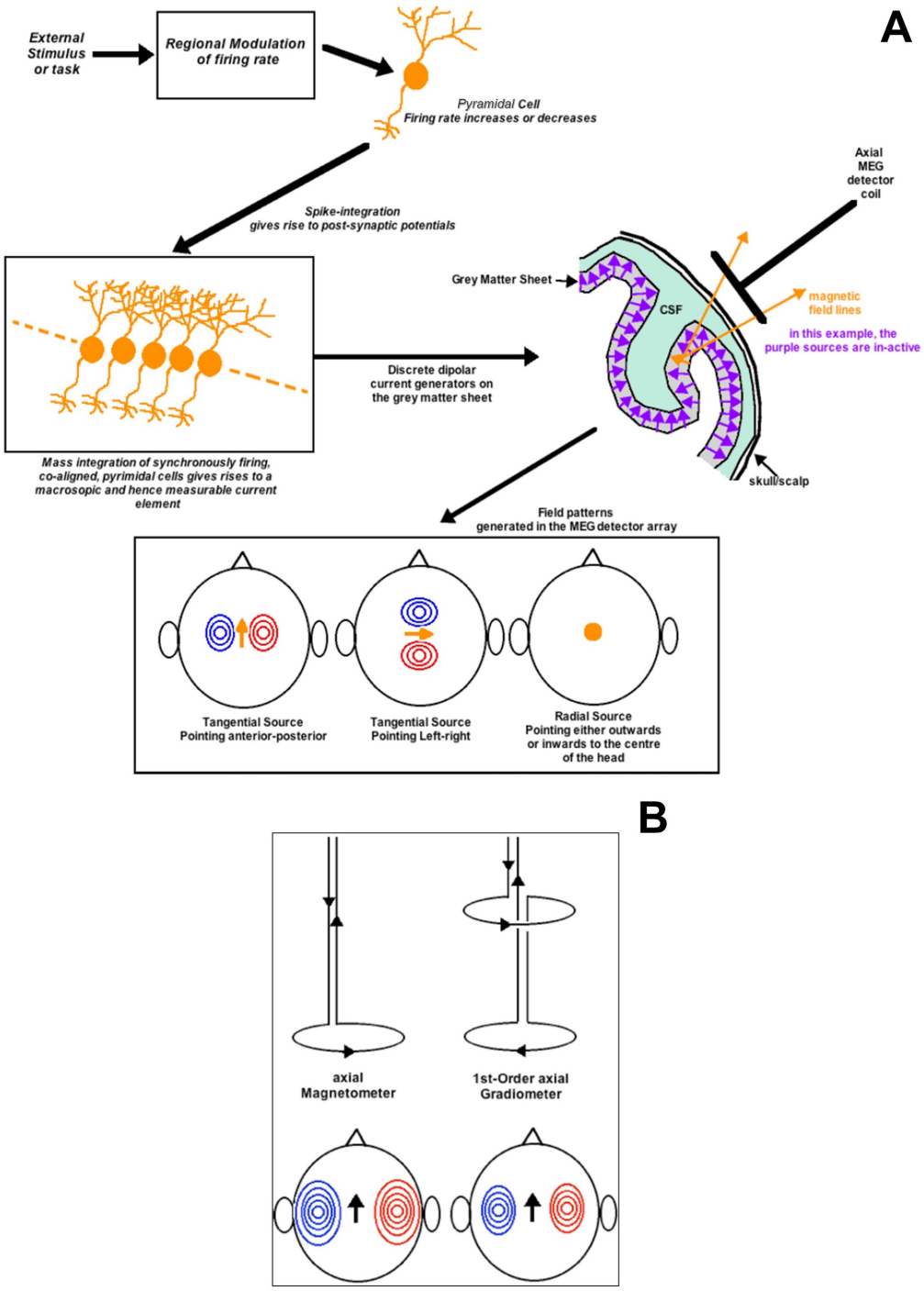

FIGURE 1.3: **A.** Generation of magnetic fields from the synchronized activity of a pyramidal neuron population. The bottom panels show source configurations as captured by MEG sensors, with red and blue lines showing magnetic fields entering and exiting the head respectively. Reproduced from Singh (2006). **B**. Field patterns generated by a tangential source using an axial magnetometer or a first-order axial gradiometer with a compensation coil. Adapted from Singh (2006).

To detect the magnetic field over a larger area and relay it to the SQUIDs, flux transformers are used, which consist of a pick-up coil (or magnetometer) and a coupling coil (Vrba and Robinson, 2001). As external magnetic noise can prevent the detection of weak neural magnetic fields, noise rejection strategies are used in modern MEG systems, starting with the design of the pick-up coils. For example, axial gradiometers use a pick-up coil together with a compensation coil wound in the opposite direction (Figure 1.3B). This design takes advantage of the spatial gradient of the magnetic field, which falls off rapidly with distance: variations in the background field are measured by both coils and effectively cancelled out, while signals of interest cause a larger change in the spatially closer pick-up coil (Hämäläinen et al., 1993). More complex combinations of coils can improve noise rejection performance (Singh, 2006). The CTF MEG system, used for MEG recordings described in this thesis, consists of 275 first-order axial gradiometers and 29 reference magnetometers, which are used to regress out additional noise in post-processing and implement synthetic third-order gradiometers (Vrba and Robinson, 2001). Furthermore, all recordings are conducted inside a magnetically shielded room (MSR) which attenuates environmental noise.

1.2.2 Strengths and challenges

Due to the different properties of electric and magnetic fields, MEG is thought to be more sensitive than EEG to primary (intracellular) currents and less affected by volume (extracellular) currents, whose magnetic fields tend to cancel out (Vrba and Robinson, 2001). Moreover, MEG, unlike EEG, does not require a reference electrode, and is less susceptible to muscle artefacts due to reduced volume conduction effects (Claus et al., 2012; Muthukumaraswamy, 2013). However, to generate measurable magnetic fields outside the head, neuronal sources must be oriented tangentially to the skull and not radially (Figure 1.3A). In practice, this is not a major limitation of MEG, as radial sources located at the crests of gyri are thought to form less than 5% of the cortical area. A more limiting factor is the lower sensitivity of MEG to deep sources, caused by the fact that magnetic fields decay rapidly with distance (Hillebrand and Barnes, 2002). Recent research, however, has shown

successful source localization of responses from deep structures such as the hippocampus (e.g. Meyer et al., 2017).

Although MEG has excellent temporal resolution, the localization of sensor-level responses can be more ambiguous. The inverse problem of MEG source localization is ill-posed (Sarvas, 1987): given a magnetic field measured by MEG, there is an infinite number of possible cortical source distributions that could have generated it. Though there are several methods of alleviating the inverse problem by imposing prior constraints, source analyses in this thesis use a linearly constrained minimum-variance (LCMV) beamforming approach (Hillebrand et al., 2005; Van Veen et al., 1997). This method independently estimates a solution at each source location in the brain by weighting the sensor-level measurements so as to increase sensitivity to the location of interest, while minimizing interference from other locations. To achieve this, a forward model specifying the sensor pattern for each active source (Mosher et al., 1999) is combined with the data covariance matrix. The LCMV approach estimates a vectorial solution comprising all three possible dipole orientations, which can be reduced to a scalar solution using Singular Value Decomposition (SVD); both approaches have been used in this thesis.

Beamforming has a few advantages: it attenuates noise (Vrba, 2002), it does not entail assumptions about the number of active sources (Robinson and Vrba, 1999), and only assumes no strong temporal correlations between sources (Hillebrand et al., 2005). Furthermore, although beamformer images can have a non-uniform spatial resolution, they have been shown to resolve active sources with a resolution between ~2-20 mm (Barnes et al., 2004).

In sum, MEG provides rich whole-head direct measurements of neural activity, with excellent temporal and spectral resolution, and good source reconstruction resolution despite inherent ambiguity. Furthermore, recent technological advances signal a bright future for MEG. While currently SQUID sensors need to be placed in a cryogenic dewar and are thus situated at a distance from the subject's head, optically-pumped magnetometers (OPMs) have been developed that can be placed directly on the scalp, with the potential to significantly increase signal-to-noise ratio (SNR) and spatial resolution in MEG (Boto et al., 2017, 2018; Iivanainen et al., 2017).

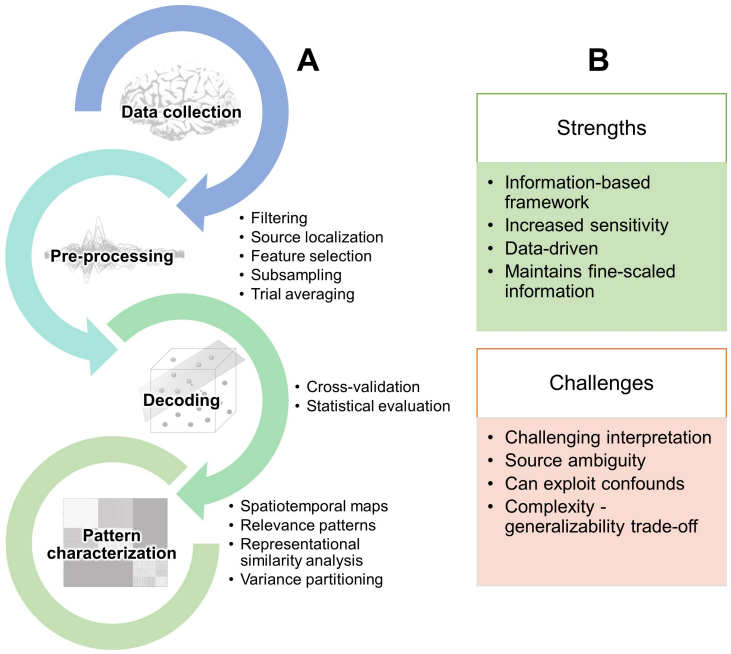

FIGURE 1.4: **A.** Multivariate analysis framework from data collection to model evaluation. Note that some of the analyses in the final step, such as RSA, can be performed independently of the others, using distance metrics other than decodability. **B**. Summary of the main strengths and challenges of MVPA analyses.

1.3 Machine learning for MEG

With high-density spatial sampling and millisecond-resolved temporal resolution, MEG captures neural activity in rich, high-dimensional datasets that can pose an analysis challenge, especially in the absence of fully standardized pipelines or prior information about the phenomenon under study. As opposed to univariate statistical methods which often rely on signal averaging, multivariate methods offer increased sensitivity by exploiting information in distributed patterns, and can help reveal underlying structure in such complex data. As such, they are increasingly being adopted in the analysis of neuroimaging data, bringing new challenges along with new insights.

Multivariate pattern analysis (MVPA) methods have been adopted from machine learning, where the focus is on training informationally greedy algorithms to obtain accurate out-of-sample predictions for real-world applications. While the prediction goal is also valid for some neuroscience applications (such as clinical data), in most cases machine learning is applied to neuroimaging data with a completely different goal: understanding the brain. This focus on interpretation

(Hebart and Baker, 2017) changes the way in which we apply, constrain, and evaluate the algorithms, and will be discussed in more detail below.

Machine learning can be defined as a set of algorithms that automatically learn to generalize from examples (Domingos, 2012). Two of the main categories of methods used in machine learning are supervised learning (in which the algorithm is provided with labelled examples during training) and unsupervised learning (in which an algorithm is used to uncover structure in unlabelled data). The latter is most commonly used in neuroimaging for visualization and dimensionality reduction.

Multivariate analysis is usually performed using classification algorithms, which constitute a subset of supervised learning methods, alongside regression. While in neuroimaging regression is often used to predict neural time series based on the design matrix, classifiers are used to predict the experimental conditions from neural patterns, thus reversing the direction of the inference (Pereira et al., 2009). More specifically, classifiers predict the class (category) of previously unseen examples (data points) based on the value of their features (e.g. sensor signal amplitudes).

To make their predictions, classifiers learn a number of parameters from a training dataset and create a model of the relationship between features and class labels. To determine if the features contain information about data classes, the trained classifier is tested on new data and the out-of-sample generalization performance is computed, most commonly in terms of accuracy (proportion correctly labelled examples in the test set). In the case of MEG data, a dataset could contain single trials as examples, and magnetic field amplitudes at all MEG sensors as features. The classifier might be trying to learn the relationship between MEG sensor patterns and the type of visual stimulus presented to the subject, e.g. face or house.

A MEG multivariate analysis pipeline typically starts with data pre-processing (Figure 1.4A). Dimensionality reduction is sometimes performed, which can entail a subselection of sensors, sources or time windows, or data-driven methods such as Principal Component Analysis (although see Goddard et al., 2017 for some caveats). The choice of features can strongly affect the interpretability of the data, including the spatiotemporal resolution and generalizability of the results.

Other steps carried out aim to increase the SNR of the data, for example by averaging subsets of trials, and to ensure balanced classes by subsampling the majority class. Although pre-processing choices can affect the performance of the decoding algorithm (Grootswagers et al., 2017), a high accuracy is not important in studies using decoding for understanding brain function rather than prediction (Hebart and Baker, 2017); rather, the presence of discriminating information is assessed through statistical testing against the chance level. Since decoding accuracy is a relative measure of effect size, preventing cross-experiment comparisons, differences in pre-processing choices across experiments (including in this thesis) can be considered less important than unbiased classifier testing, evaluation, and statistical assessment.

Although decoding can be performed with a variety of algorithms, linear classifiers are most commonly used in neuroimaging data analysis as a linear readout of the data is biologically plausible and offers increased interpretability compared to more complex algorithms (DeWit et al., 2016; Kriegeskorte and Kievit, 2013; Ritchie and Carlson, 2016). Note that in cases in which the choice of algorithm or its hyperparameters need to be optimized for maximal prediction accuracy, this optimization should be performed on a third independent subset of the data (validation set; Lemm et al., 2011). However, for many neuroimaging applications, linear classifiers with default hyperparameters are sufficiently powerful to reveal the presence of decodable information.

1.3.1 The Support Vector Machine classifier

Throughout this thesis, a linear Support Vector Machine (SVM) classifier (Boser et al., 1992) is used for binary decoding of MEG data. For datasets composed of n examples x_i with label $y_i \in \{-1, 1\}$, a linear classifier is based on a linear discriminant function

$$f(x) = \mathbf{w}^T\mathbf{x} + b \tag{1.1}$$

where the dot product $\mathbf{w}^T\mathbf{x} = \sum_i w_i x_i$, $\mathbf{w}$ is known as the weight vector, and b is the bias. The sign of the discriminant function divides the dataset using the

set of points $\mathbf{x}$ such that $\mathbf{w}^T\mathbf{x} + b = 0$, which form a line in a 2D space, a plane in a 3D space, and a *hyperplane* in higher dimensional spaces (Ben-Hur and Weston, 2010). This is known as the decision boundary and is also the basis of a linear SVM (Figure 1.5).

A linear SVM finds the maximal margin hyperplane between two classes by maximizing the distance between the decision boundary and the data points and thus increasing its generalizability. For data that are not perfectly linearly separable, a soft-margin SVM is used, which allows the misclassification of some examples by introducing slack variables ξ_i. The trade-off between error rate and margin size is controlled by a regularization parameter, the box constraint C, which penalizes misclassified examples (Noble, 2006).

The problem of maximizing the geometric margin $1/\|\mathbf{w}\|$ is equivalent to minimizing $\|\mathbf{w}\|^2$. The optimization problem solved by SVMs in what is known as the *primal* formulation is thus

$$\underset{w,b}{\text{minimize}}\ \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_i \xi_i \tag{1.2}$$

This is known as an L1-SVM (which imposes a linear loss for margin-violating examples), while an L2-SVM imposes a quadratic loss (a larger penalty), and differs only in the regularization term, which is $\frac{C}{2}\sum_i \xi_i$. Both types of regularization have been used in the analyses presented in this thesis.

Solving the primal optimization problem for large datasets would be computationally prohibitive, especially when mapping the data onto a higher-dimensional space. SVM implements a sparse and more tractable solution by selecting a subset of the examples x_i located closest to the hyperplane, known as *support vectors*, for whom the Lagrange multipliers $\alpha_i > 0$. This is known as the *dual* formulation, in which the optimization problem becomes

$$\begin{aligned}\underset{\alpha}{\text{maximize}} \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} y_i y_j \alpha_i \alpha_j (x_i^T x_j),\\ s.t. \sum_{i=1}^{n} y_i \alpha_i = 0, 0 \le \alpha_i \le C\end{aligned} \tag{1.3}$$

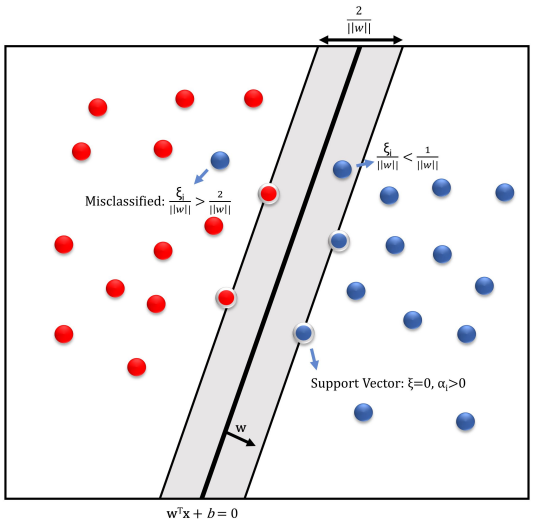

FIGURE 1.5: Visualization of a soft-margin SVM applied to non-linearly separable data in 2D. The classifier relies on support vectors to maximize the margin, and includes a regularization parameter that penalizes misclassified data points.

and the discriminant function becomes

$$f(x) = \sum_{i=1}^{n} \alpha_i y_i (x_i^T \mathbf{x}) + b. \tag{1.4}$$

Note that in non-linear classifiers, the dot product $x_i^T \mathbf{x}$ is replaced by a different kernel function. For linear SVM, the weight vector can be recovered based on the input examples:

$$\mathbf{w} = \sum_{i=1}^{n} y_i \alpha_i x_i \tag{1.5}$$

The in-built regularization and efficient handling of a large feature space (Nilsson et al., 2006) make SVM a good approach for high-dimensional neuroimaging datasets, and the weights associated with each feature, although not directly informative, can help uncover the spatial patterns underlying successful classification.

1.3.2 Cross-validation and statistical evaluation

To measure the prediction performance of a classifier, the trained model needs to be tested on an independent dataset. Although this can be done by holding out part of the data for testing, such a procedure does not exploit the full dataset, which can be an issue given the small sample sizes common in neuroimaging. A more

commonly used method is cross-validation, which involves splitting the dataset into partitions (folds) and holding out each one of them for testing, while training on the remaining data. The final accuracy is averaged across all folds.

Cross-validation involves a trade-off between bias and variance (underfitting and overfitting), depending on the number of partitions. In leave-one-out cross-validation, each example is used for testing a model trained on all other examples; although this type of cross-validation is exhaustive, it is computationally expensive and can lead to variable estimates (Lemm et al., 2011). In stratified k-fold cross-validation, the data are randomly split into k folds (commonly 10 or 5), ensuring balanced class representation in each fold, and classification accuracy is averaged across folds. This can be a more efficient alternative (Pereira et al., 2009), and is the method used for calculating classification accuracy in this thesis (Figure 1.6).

When using MVPA to make inferences about the brain, it is important to assess the presence of decodable information against the null hypothesis, which predicts chance-level classification performance. Throughout this thesis, this is done through randomization testing. The estimation of an empirical null distribution is important given the often skewed distributions of accuracies in small datasets (Jamalabadi et al., 2016), and has been shown to assess significance more reliably than theoretical chance levels (Combrisson and Jerbi, 2015) or binomial tests (Noirhomme et al., 2014). In the analyses that follow, label shuffling across training and test sets was used to estimate null accuracy distributions and calculate p-values (Figure 1.6).

1.3.3 Resolving temporal dynamics

In MEG MVPA studies and throughout this thesis, decoding is usually performed in a time-resolved manner, allowing discriminating information about the experimental conditions to be detected with high temporal accuracy and often earlier than the typical evoked responses (Cichy et al., 2015; Grootswagers et al., 2017). An alternative approach is cross-decoding across time points, by using each time point for training a separate model and testing it on held-out data from every other time point (King and Dehaene, 2014).

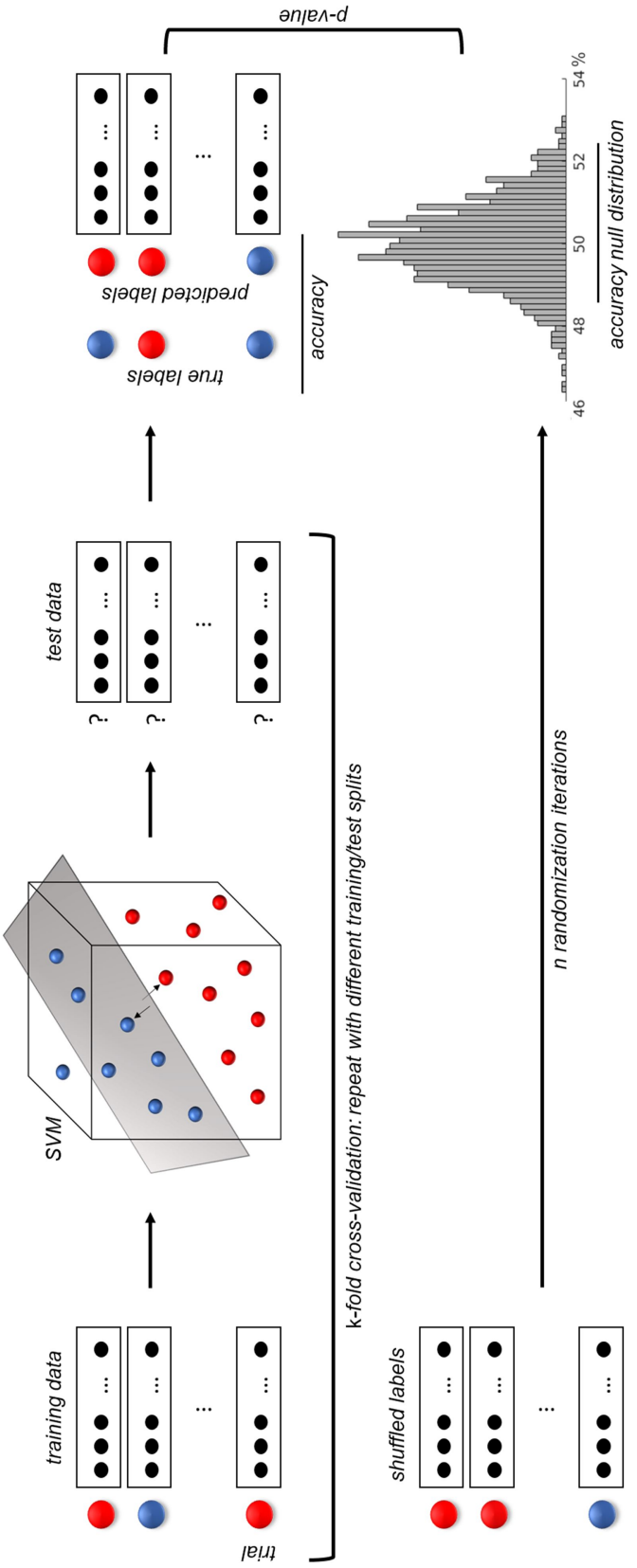

FIGURE 1.6: Overview of the decoding analysis steps used throughout this thesis. Classification accuracy is calculated using k-fold cross-validation and statistical significance is assessed through randomization testing.

This temporal generalization approach evaluates the temporal structure of neural representations: if responses are sustained, classifier models are expected to generalize over time, while transient responses are characterized by rapidly changing classifier weights. It is thought that stable representations are associated with conscious perception and recurrent processing (Dehaene, 2016; Mohsenzadeh et al., 2018b), although trial-to-trial variability has been suggested as a potential alternative explanation for the quick succession of temporal stages revealed by this method (Vidaurre et al., 2018). Although the interpretability of resulting accuracies depends on experimental design and potential confounds, this method exploits the temporal resolution of MEG and sensitivity of MVPA to offer a putative link between brain mechanisms and perceptual processes (Chapter 3).

1.3.4 Uncovering spatial information

One of the main challenges in MVPA is source ambiguity, or recovering the spatial patterns leading to successful decoding (Carlson et al., 2017; Naselaris, 2015; Tong and Pratte, 2012). Although several approaches have been proposed and are explored in the experimental chapters of this thesis, they entail different assumptions and pose interpretation challenges.

A main issue in exploring spatial correlates is choosing the right spatial scale. In a whole-brain approach, large-scale distributed patterns may be exploited, which can render the method more powerful; on the other hand, we may wonder if such large-scale information from disparate regions can be used by the brain, or is solely available to the experimenter (Carlson et al., 2017). Furthermore, whole-brain analyses often suffer from the "curse of dimensionality" (Scott, 1992). The converse approach of decoding from regions of interest or searchlights (uniform patches across the brain) entails the assumption that information is represented locally (Kragel et al., 2018). Furthermore, information can sometimes be combined or segregated suboptimally in such analyses, and the multiple tests conducted can also pose a concern (Tong and Pratte, 2012). Comparing models at different spatial scales can help resolve these differences, and prior information can elucidate source ambiguities (Carlson et al., 2017).

In MEG MVPA studies, most decoding analyses are implemented at the sensor level, with few studies performing source-space decoding (e.g. Su et al., 2012). Although sensor-space signals can be informative, they are less likely to be consistent between participants and more prone to signal leakage (Zhang et al., 2016b). On the other hand, source-space methods can decrease classification performance, while also suffering from concerns of signal leakage and information spreading (Gohel et al., 2018; Sato et al., 2018). Throughout this thesis, sensor-space classification is employed as a benchmark for temporal information, while source-space classification is used to investigate the spatial dynamics of decodable information.

Whole-brain approaches rely on feature weights to assess the contribution of different sensors or sources to the decoding performance. However, classifiers can exploit non-informative features in generating predictions, and their weights are thus not directly interpretable. A procedure has been proposed to recover activation patterns from feature weights using the data covariance matrix (Haufe et al., 2014), and this solution is implemented in Chapter 2. One main caveat when interpreting weight-based maps is that inferences can only be made about a feature relative to the others, since the weights are specific to the feature set used in decoding (Williams and Henson, 2018). To overcome this concern, the analysis in Chapter 2 uses a dimensionality reduction method that creates unique and equally weighted features for each of 84 ROIs across the brain; thus, the contribution of each ROI can be evaluated using whole-brain relevance maps.

Other methods of mapping classification accuracy involve spatial selection of sources. Searchlight methods (Kriegeskorte et al., 2006) originating in fMRI have become widely used, due to their high spatial resolution and hypothesis-free coverage of the whole brain. Although volumetric searchlights can inaccurately represent information as being uniformly distributed in the brain (Etzel et al., 2013), this is less of a concern in MEG, where spatial maps do not have the resolution of fMRI. For example, in Chapter 3, source activity is reconstructed using a 10 mm grid and searchlight analysis is performed using clusters of neighbouring sources; the additional smoothing introduced by the searchlight is not likely to be problematic given the spatial resolution of MEG. However, MEG searchlight maps need to be interpreted cautiously given the source ambiguity, spatial smoothing, and signal

leakage concerns that can be compounded by the use of sensitive algorithms.

An alternative to searchlight approaches is decoding from functional ROIs (Chapter 5). This can be less computationally expensive and allow for better cross-subject and cross-study integration, as well as improving interpretability (Hillebrand et al., 2012). However, different ROI sizes and potential SNR differences can make comparisons across regions difficult (Haynes, 2015).

Although source localization of information maps is challenging, similar interpretation and reliability concerns are also inherent in univariate MEG source analyses, as well as other neuroimaging methods. Combined with a sensor-level benchmark for the presence and temporal dynamics of an effect, the source-space decoding analyses in this thesis contributed complementary information, suggesting that when cross-modal investigations are not possible, MEG can offer a rich picture of the spatiotemporal dynamics of high-level vision. Although source-space decoding methods are still in their early stages, the range exemplified here suggests that different questions can be answered using different approaches, depending on prior information and hypotheses. Finally, the source localization capabilities of MEG MVPA are likely to rapidly improve given recent advances in machine learning algorithms, together with a growing understanding of the challenges and caveats of these methods in the context of neuroimaging, and technological advances such as on-scalp MEG.

1.3.5 Characterizing patterns: Representational Similarity Analysis

Another difficulty in interpreting MVPA results lies in their representational ambiguity (Carlson et al., 2017; Naselaris, 2015), or the difficulty of understanding how decodable information is represented in the brain. Explicit modeling approaches can be employed alongside or instead of MVPA to tease apart the content of brain representations (Naselaris, 2015; Poldrack, 2011). Although the concept of brain representation is in itself ambiguous, it has been defined as a latent variable expressing shared variance between brain activity and outcome measures (Kragel et al., 2018). Thus, investigations of representational structure work at a level with the potential to link psychological constructs to neural substrates (Ritchie et al., 2017).

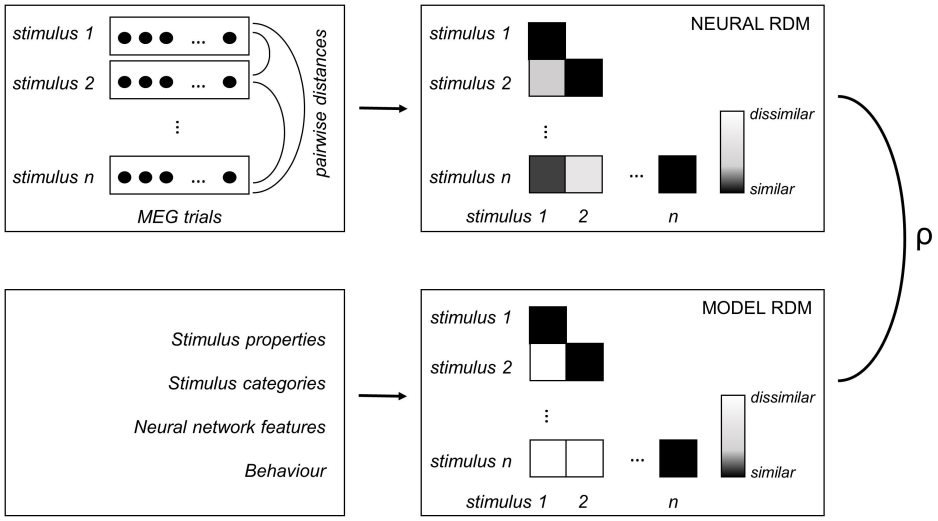

FIGURE 1.7: Representational Similarity Analysis. Pairwise distances between stimulus responses are calculated in order to create neural representational dissimilarity matrices (RDM), which are compared to model dissimilarity matrices using Spearman's rank correlation.

A popular approach for investigating the content of neural patterns is representational similarity analysis (RSA; Kriegeskorte, 2011; Kriegeskorte and Kievit, 2013; Kriegeskorte et al., 2008). The main appeal of this method is its ability to bring together measures from different modalities within a common representational space in order to search for shared structure. Neural patterns from different modalities can thus be combined and compared to models based on behaviour, physiology, stimulus properties, machine learning, or theory. The variance explained by each model can be quantified, compared to other models and evaluated against a noise ceiling specifying the maximal possible performance (Nili et al., 2014).

RSA starts with the choice of a distance metric to capture the similarity structure in the data. Different metrics have been used, including Euclidean and Mahalanobis distances, decoding accuracies, and correlation distances, with recent evaluations suggesting that cross-validated distances are the most reliable in the presence of noise (Guggenmos et al., 2018). The choice of metric can impact how well the underlying similarity structure is captured (Carlson et al., 2017). In MEG RSA, the distance metric can be applied to trials corresponding to all pairs of stimulus exemplars in order to obtain a neural representational dissimilarity matrix (RDM; Figure 1.7). Next, model dissimilarity matrices are created quantifying the

predicted pairwise distances between stimuli based on different hypotheses. For example, if neural data encoded contrast, we might expect the neural RDM to correlate highly with a model RDM quantifying differences in contrast between stimuli; while a high-level representation might correlate better with a binary model dividing the stimulus set along category axes. Model RDMs are compared to the neural data using a rank correlation, since a linear relationship cannot usually be assumed when using non-invasive measures of neural patterns (Nili et al., 2014).

Although a simple correlation metric is used to assess the relationship between complex representational spaces (Carlson et al., 2017), additional steps can be performed to maximize the interpretability of RSA results. First, a noise ceiling can be calculated to quantify the variance explained by the true model given the noise in the data (Nili et al., 2014). To calculate a lower bound, subject-wise neural RDMs are correlated to the average neural RDM across the remaining subjects, and an average correlation coefficient is obtained using a leave-one-out procedure. Next, subject-wise neural RDMs are correlated to the average neural RDM across all subjects to obtain an upper bound of the noise ceiling. Since the former estimate underfits the true correlation, while the latter overfits, the true model correlation is expected to fall between the two bounds. Next, partial correlations can be used to quantify the unique variance explained by models of interest after removing confounding models (e.g. Bonner and Epstein, 2017; Chapter 3, Chapter 5), and variance partitioning can help visualize the shared and unique variance contributed by a group of models (e.g. Groen et al., 2018; Chapter 3).

Like decoding analyses, MEG RSA can also be performed with varying spatiotemporal resolutions. While neural RDMs are often computed from the whole-head MEG sensor patterns (e.g. Pantazis et al., 2017; Wardle et al., 2016), in this thesis, space-varying RDMs are used to explore the progression of feature representations across the visual system. In line with the MVPA analyses, both searchlight (Chapter 3) and functional ROI (Chapter 5) RSA mapping was performed. All RSA analyses were conducted in source space, maximizing inter-subject correspondence for a fixed-effects procedure, and statistical evaluation was performed using randomization testing.

1.3.6 Challenges in multivariate analysis

As multivariate methods for neuroimaging have increased in sophistication, there has been a growing awareness of the caveats and challenges in their implementation (Figure 1.4B). These stem mainly from the transition between an activation-based and an information-based framework, leading to difficulties in interpreting results and eliminating confounds.

Decoding methods can tease apart distributed and overlapping patterns, and exploit fine-scale information rather than averaging it. Despite the source ambiguity discussed above, this leads to increased sensitivity in detecting effects (Haynes and Rees, 2006; Tong and Pratte, 2012; Varoquaux and Thirion, 2014; Williams and Henson, 2018). However, this sensitivity has two alternative explanations that are not linked to neural activity. The first one is the switch from activation-based, directional tests to information-based tests that discard the direction of an effect and only quantify the presence of information (Friston, 2009; Hebart and Baker, 2017; Varoquaux and Thirion, 2014). Although this is not an issue in itself, group analyses that average non-directional information metrics can speak only to the availability of discriminating information, unlike univariate analyses focusing on signal increases. Furthermore, for experiments performing within-subject decoding (as described in this thesis), it is important that potential confounding variables are controlled at the subject level and not just at the group level, in order to avoid spurious group effects (Todd et al., 2013).

The second potential explanation of an increased sensitivity is the contribution of confounds to decoding performance. An often-cited example of this is the outcome of the 2006 Pittsburgh Brain Competition (Tong and Pratte, 2012), where a team achieved successful decoding of humorous scenes from fMRI signal in the ventricles, likely due to stimulus-correlated head motion. Other types of confounds can increase prediction accuracy, such as low-level differences in stimulus properties (Cox and Savoy, 2003). Since decoding accuracies can reflect differences in variability (noise) as well as means (signal), it is important to ensure that the variability does not reflect unrelated confounds (Hebart and Baker, 2017). These concerns can

be alleviated by using controlled stimulus sets (where stimuli are matched along irrelevant dimensions or confounding properties are orthogonal to the properties of interest; Bracci et al., 2017a), or by demonstrating cross-exemplar or cross-category generalization (Kragel et al., 2018; Tong and Pratte, 2012). Confounding properties can also be explicitly modelled in analyses like RSA, allowing them to be removed from the analysis. All three of these approaches are used in the experimental chapters of this thesis to maximize the interpretability of decoding and RSA results.

The final (and desired) source of increased MVPA sensitivity is the ability to exploit multivariate patterns, including their covariance structure, in agreement with a view of the brain as an information processing system employing population coding (Ritchie et al., 2017). This leads us to the next challenge in MVPA analysis, which is related to the interpretability of decoding results.

The biological plausibility of a linear readout of population codes (DeWit et al., 2016; Kriegeskorte and Kievit, 2013) has led to assumptions that these representations are used by the brain, although this cannot be directly shown by MVPA analyses (Carlson et al., 2017; Ritchie et al., 2017; Yamins and DiCarlo, 2016). Like other neuroimaging methods, decoding is inherently correlational (Jonas and Kording, 2017; Poldrack, 2011). Avoiding non-linear transformations of the data or the use of information likely to be inaccessible to the brain (i.e. combinations across disparate regions) can improve plausibility, but cannot establish causality. Similarly, the absence of decodable information cannot be interpreted as absence of information within the neural population, since the relevant information could be organized in ways inaccessible to decoding algorithms (Haynes, 2015).

To increase interpretability, it is often suggested that decoding results should be linked to behaviour, as not all decodable information contributes to behavioural responses (Grootswagers et al., 2018; Williams et al., 2007). In order to connect neural representational spaces to psychological constructs, neural patterns can be used to predict behaviour, for example within a RSA framework (Carlson et al., 2017; Ritchie et al., 2017; Chapter 3).

An additional challenge in MVPA is navigating the trade-off between model complexity and generalizability. Although linear classifiers are popular in neuroimaging due to the reasons described above, the recent success of deep neural

networks (DNNs) offers an alternative computational model. Trading off accuracy with complexity, DNNs have achieved near-human performance on object recognition tasks (Kietzmann et al., 2017) through increasingly complex architectures (e.g. Krizhevsky et al., 2012). In understanding the brain, simpler models provide more knowledge and are more explicit than complex models (Turner et al., 2018); for example, complex operations like the ones performed by DNNs could achieve stimulus category predictions based on retinal patterns, in the absence of explicit representations of category in the data (Kragel et al., 2018). On the other hand, DNN layer activations have shown striking similarities with the human visual system (Cichy et al., 2016; Groen et al., 2018; Yamins and DiCarlo, 2016; Chapter 5), and their "black box" quality has been reduced by investigations of the features they use to achieve category representations (e.g. Bonner and Epstein, 2018).

Opinions on the role of DNNs in cognitive neuroscience span a broad range between viewing them as potential models with certain constraints (Scholte, 2018; Turner et al., 2018) and viewing the brain itself as a DNN system, which internally optimizes cost functions for specific problems (Marblestone et al., 2016). Although part of this can be ascribed to a tendency to liken the brain to the computational advance of the day, it is certain that DNNs have much to contribute as models that can be optimized for biological plausibility and trained on specific tasks, thus potentially overcoming the complexity challenge.

1.4 Investigating face and scene perception

Although the neural correlates of face perception have been reliably mapped, it is still not well understood how the visual system efficiently represents information, allowing us to recognize people and discern social cues at a glance. Similarly, we effortlessly understand and navigate our environment, but there is significant debate around the neural computations underpinning this ability. To address these questions, the experiments in this thesis approach emotional face and natural scene perception with the multivariate analysis tools described above.

Since the discovery of face-selective cells and brain areas (Gross, 2002; Kanwisher et al., 1997), the study of face perception has been marked by debate about

the organization of neural face processing systems: modular or distributed, horizontal or hierarchical (Haxby et al., 2001; Kanwisher, 2000). Face perception is supported by a cortical network in the ventral visual cortex (Ishai, 2008), which includes the occipital face area (OFA), fusiform face area (FFA), and the superior temporal sulcus (STS). Although these regions show reliable signal increases in response to faces, the type of face information they represent is still the subject of debate. A classic model (Bruce and Young, 1986) proposed a template-matching view of facial recognition, whereby several types of information are extracted from faces and compared to stored structural codes. The suggestion made here, that identity and expression information are separately extracted, was further developed within a neuroanatomical framework (Haxby et al., 2000). This model entailed a core system (extrastriate visual areas and OFA) relaying changeable face information (including expression) to the STS, and invariant face features (including identity) to the FFA. This core network was thought to communicate with an extended system consisting of subcortical, parietal and anterior temporal structures.

However, other studies have shown that expression and identity are integrated at an early stage (Calder and Young, 2005) or that expression is processed in the FFA (Bernstein and Yovel, 2015). Thus, an alternative model suggests that the dissociation between form and motion processing is what drives the distinction between the two pathways (Duchaine and Yovel, 2015; Pitcher et al., 2011). However, evidence of parallel connections within the face network (Pyles et al., 2013) and of interaction between the two streams (Fisher et al., 2016) suggests that the two pathways are not functionally segregated. Furthermore, evidence of increasingly invariant identity representations along the ventral stream point to a feature processing hierarchy, similar to models discussed in section 1.1. On the other hand, information appears to be integrated both locally and within larger-scale networks, suggesting that both modular and distributed codes support face perception (Freiwald et al., 2016).

Efficient face processing is thought to be supported by coarse, feature-based face detection followed by configural processing (Calder et al., 2000; Maurer et al., 2002; Piepers and Robbins, 2012). This is associated with a holistic representation

(Farah et al., 1998; Richler and Gauthier, 2014) which leads to much lower performance in extracting information from inverted faces (Behrmann et al., 2014; Yin, 1969). It is thought that faces may be represented as points in a high-dimensional "face space" based on feature axes (Leopold et al., 2001), and combining such a model with electrophysiology and multivariate analysis has recently led to the recovery of a low-dimensional identity code in primates (Chang and Tsao, 2017). The axes along which face features are represented could be matched to patterns in face-selective areas, which may represent faces featurally or topologically (Henriksson et al., 2015). However, it is possible that different codes or "face space" axes govern the representation of different face dimensions, such as identity or expression, and that these may vary according to task effects. Differences in cytoarchitecture between face-selective regions suggest that different computations may be performed within each region (Grill-Spector et al., 2018); it is thus possible that efficient face processing relies on sparse featural representations implemented in a modular fashion and rapidly accessible to distributed, large-scale systems.

The first three chapters of this thesis focus on emotional face perception. Emotional cues are highly salient and recruit distinct systems in the face processing network, with a putative direct subcortical thalamus-amygdala route thought to rapidly relay coarse face information (Vuilleumier et al., 2003;Figure 1.8). However, it is unclear whether this pathway is emotion-specific (Garrido et al., 2012; Garvert et al., 2014; McFadyen et al., 2017), or whether rapid expression perception is instead supported by rapid cortico-cortical loops (Liu and Ioannides, 2010; Pessoa and Adolphs, 2010; Pourtois et al., 2013).

Although multivariate analyses have demonstrated rapid encoding of face identity (Davidesco et al., 2014; Nemrodov et al., 2016; Vida et al., 2017), expression has been investigated to a lesser extent with such approaches (but see Cecotti et al., 2017; Tsuchiya et al., 2008; Wegrzyn et al., 2015; Zhang et al., 2016a). In this thesis, expression processing is explored using different task contexts, controlled stimulus sets, and MEG MVPA analyses. Multivariate approaches can help resolve disagreements on the timing of expression processing, as well as investigate its spatiotemporal dynamics in a single, data-driven framework. Furthermore, model-based approaches like RSA can directly test opposing hypotheses within a single

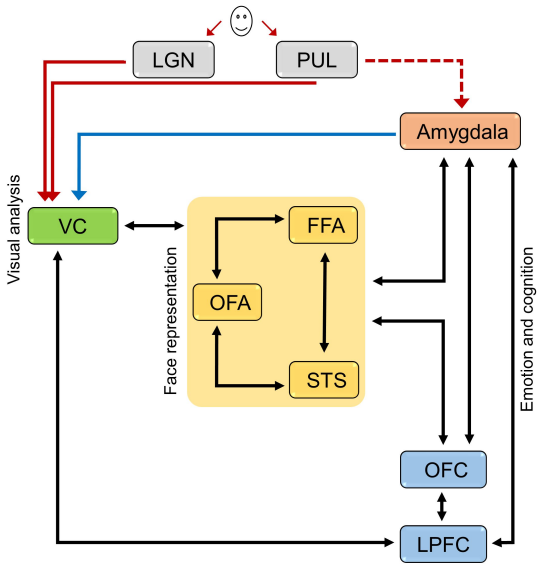

FIGURE 1.8: Simplified visualization of the network relaying visual affective information from faces, loosely based on Figure 2 from Pessoa, 2008. Red, blue and black arrows show feedforward, feedback and reciprocal connections respectively. The dashed line shows the putative direct subcortical route to the amygdala. For simplicity, not all connections are shown. LGN: lateral geniculate nucleus; LPFC: lateral prefrontal cortex; OFC: orbitofrontal cortex; PUL: pulvinar;VC: visual cortex.

dataset in a spatiotemporally resolved manner, thus offering potential explanations for previous conflicting findings (Chapter 3). Given the success of computer vision in achieving object recognition and its growing role in cognitive neuroscience (VanRullen, 2017), it is likely that computational models for increasingly complex and naturalistic tasks will become available. Face processing as implemented in the ventral stream may both inform (Grill-Spector et al., 2018) and be informed by such models, which could finally link psychological models of face perception with representational axes in the brain.

At this point, it is important to mention a caveat to the approach to expression perception described in this thesis: although a vision-focused approach can help uncover the computations transforming salient low-level features into expression representations, it is important to remember that such processes are part of larger systems involved in emotion and social cognition. Using highly controlled, static stimuli (as in all experiments reported here) can help isolate the phenomena of interest (the extraction of salient visual cues), but not their social component

(Teufel et al., 2013). Although here face perception is considered from an information processing perspective, future studies can employ more complex stimulus and experimental designs in order to compare social settings for expression processing to purely visual settings.

The final chapter of this thesis investigates natural scene perception (Chapter 5). Although not as salient as faces, scenes are ubiquitous in our daily life and we can effortlessly extract their "gist" (Rousselet et al., 2005). A visual network has been shown to preferentially respond to scenes (Dilks et al., 2013; Epstein, 2008; Epstein and Kanwisher, 1998), including the parahippocampal place area (PPA), the retrosplenial cortex (RSC), and the occipital place area (OPA). However, the representational content of these areas remains the subject of debate, as contradictory findings have shown them to encode low-level features (Nasr et al., 2011; Rajimehr et al., 2011) or categorical dimensions (Schindler and Bartels, 2016; Walther et al., 2009). In Chapter 5, a natural scene set varying along both low-level and high-level axes is employed, and RSA analysis reveals temporally overlapping featural and categorical representations in MEG patterns.

1.5 Aims of the thesis

Bringing together machine learning approaches and MEG recordings, this thesis investigates high-level visual perception, specifically expression and scene perception. Exploring information and representation, rather than activation, is particularly suitable in questions related to high-level vision (section 1.1), where we might expect to find a link between psychological or behavioural representations and neural population coding. To maximize interpretability, information patterns are explored in space and time in order to investigate how temporal and representational dynamics might change under different task conditions.

In the first three chapters, expression processing was addressed using controlled stimulus sets containing happy, angry, and neutral faces, with different experimental paradigms. In Chapter 2, participants passively viewed emotional faces while performing a fixation cross colour change detection task. The spatiotemporal

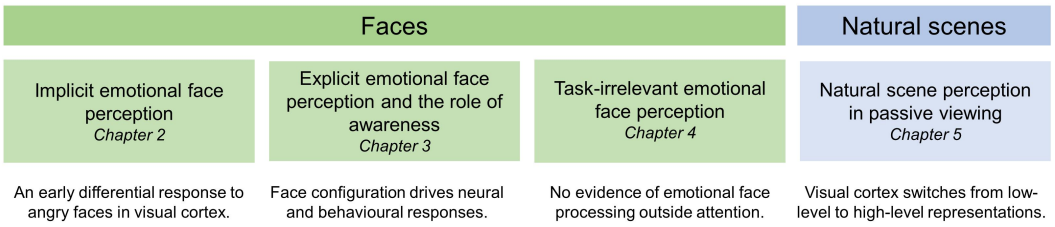

FIGURE 1.9: Thesis summary and main findings.

dynamics of expression processing were explored using sensor-level and source-space decoding, revealing whole-brain time-resolved relevance maps. The results of this experiment have been published as a peer-reviewed publication (Dima et al., 2018b). In Chapter 3, participants performed a challenging expression discrimination task with briefly presented targets, some of which were presented outside awareness. Cross-exemplar and cross-time decoding was used to investigate temporal dynamics, and RSA was performed to assess ventral representations and their link to behaviour. The results of this chapter are available as a pre-print (Dima and Singh, 2018) and have been submitted for publication. Finally, in Chapter 4, emotional faces were presented as distractors during a covert spatial attention task involving orientation discrimination. Expression processing outside attention was assessed using both univariate and multivariate analyses of electrophysiological components, broadband signals, and oscillatory activity.

The final chapter (Chapter 5) investigated natural scene perception using a passive viewing experimental paradigm identical to the one employed in Chapter 2. The stimuli were natural and urban scenes filtered at two different spatial frequencies or unfiltered. Using cross-decoding and RSA, representations of low-level features and high-level categories were assessed in sensor and source space. The results of this chapter have been published as a peer-reviewed publication (Dima et al., 2018a).

Chapter 2

Emotional faces are differentiated early in visual cortex

2.1 Abstract

Emotional faces are highly salient and are efficiently processed, but existing studies do not paint a consistent picture of the neural dynamics supporting this task. In this chapter, we addressed this question by recording MEG data while participants passively viewed a controlled set of emotional expressions and scrambled stimuli. Using time-resolved decoding of sensor-level data, we show that responses to angry faces can be discriminated from happy and neutral faces as early as 90 ms after stimulus onset and only 10 ms later than faces can be discriminated from scrambled stimuli, even in the absence of differences in evoked responses. Time-resolved relevance patterns in source space track expression-related information from the visual cortex (100 ms) to higher-level temporal and frontal areas (200–500 ms). This highlights a system optimised for rapid processing of emotional faces and preferentially tuned to threat, consistent with the important evolutionary role played by the rapid recognition of emotional cues. Furthermore, these results demonstrate that the spatiotemporal dynamics of face perception can be efficiently resolved by combining an information mapping approach with MEG sensor and source-level analyses.

2.2 Introduction

From an evolutionary perspective, it is easy to imagine why faces have a special place in the visual system, and why expression may be a particularly relevant feature to extract from other people's faces. Accordingly, the rapid extraction of emotional cues from faces is well-documented (Pessoa and Adolphs, 2010; Vuilleumier, 2005). A particular advantage seems to be afforded to threat-related expressions of fear and anger. Evidence from behavioural studies (Fox et al., 2000, 2002; Öhman et al., 2001) and neuroimaging (Feldmann-Wüstefeld et al., 2011; Pichon et al., 2012; Schupp et al., 2004) converges on the efficiency of threat detection, although the degree of automaticity with which this is accomplished is still the subject of debate (Koster et al., 2007; Mothes-Lasch et al., 2011; Pessoa, 2005).

The first part of this thesis discusses three experiments that approach emotional face perception from different perspectives (Chapter 1). In this first chapter, we focus on face perception under passive viewing and compare the results of an univariate evoked response analysis with a multivariate machine learning-based approach. We investigate whether emotional faces are decodable from MEG patterns in the absence of task-specific processing, and we explore the spatiotemporal dynamics of such an effect using an automated, whole-brain framework which requires limited prior assumptions. This information mapping approach is potentially more statistically powerful than univariate methods, and can thus help elucidate previous inconsistencies in electrophysiological research on evoked responses to faces.

The neural mechanisms underpinning rapid expression perception are still not well understood, as discussed in Chapter 1. Models postulating distinct expression and identity pathways (Haxby et al., 2000) have been challenged by evidence of expression processing in the FFA (Bernstein and Yovel, 2015), suggesting that information is extracted from faces by distributed and interacting modules (Duchaine and Yovel, 2015). A fast subcortical thalamus-amygdala route bypassing the visual cortex is thought to transmit coarse face-related information (LeDoux and Brown, 2017; Morris et al., 1998), but its role in face perception is controversial (Krolak-Salmon et al., 2004; Pessoa and Adolphs, 2011), including whether it is fear-specific

(Méndez-Bértolo et al., 2016) or non-specific to expression (Garvert et al., 2014; McFadyen et al., 2017). On the other hand, multiple fast cortical pathways forming part of a feedforward and feedback mechanism consistute an equally plausible mechanism for rapid expression perception (Liu and Ioannides, 2010; Pessoa and Adolphs, 2010).

Furthermore, electrophysiological investigations of emotional face processing in humans are not always in agreement on the temporal dynamics of expression perception. Emotional modulations of the posterior P1 evoked response component (~100 ms) are sometimes reported (Aguado et al., 2012; Eger et al., 2003; Halgren et al., 2000; Pourtois et al., 2005), with other studies failing to find early effects (Balconi and Pozzoli, 2003; Frühholz et al., 2011; Krolak-Salmon et al., 2001; Schupp et al., 2004). On the other hand, modulations of the N170 face-responsive component (120-200 ms) are consistently reported (see Hinojosa et al., 2015 for a meta-analysis).

These results point to relatively late effects, rather than the rapid differentiation of expressions which would be expected based on their preferential detection. Furthermore, categorization of other stimulus types has been detected relatively early in the visual system using multivariate methods (Cauchoix et al., 2014; Davidesco et al., 2014; Liu et al., 2009; Nemrodov et al., 2016; Vida et al., 2017). In this chapter, we aimed to look beyond ERPs, using the multivariate methods discussed in Chapter 1 to assess pattern differences in the processing of passively viewed emotional faces.

Task demands and expectations can bias visual perception (Gilbert and Sigman, 2007; Kok et al., 2012), and differences have been shown between explicit and implicit expression processing (Frühholz et al., 2011; Krolak-Salmon et al., 2001; Lange et al., 2003). Here, we opted for a passive viewing paradigm with clearly presented stimuli in order to investigate emotional face perception in the absence of an explicit task.

Using MVPA, we first interrogated the temporal dynamics underpinning expression perception, including discrimination between emotional and neutral expressions and between different emotions. Next, we applied a novel approach to source-space decoding to track the brain regions encoding the emotional content

of faces and their relative contribution over time. We were thus able to identify early differences between responses to angry faces and happy/neutral faces within 100 ms of stimulus onset and we localized them to the visual cortex, while later responses originated in higher-level temporal and frontal cortices. Our results suggest that the perceptual bias towards threatening expressions begins with the early stage of visual processing, despite a lack of significant differences in trial-averaged event-related fields (ERFs).

2.3 Materials and Methods

2.3.1 Participants

The participants were 15 healthy volunteers (8 females, mean age 28, SD 7.63) with normal or corrected-to-normal vision. All volunteers gave informed written consent to participate in the study in accordance with The Code of Ethics of the World Medical Association (Declaration of Helsinki). All procedures were approved by the ethics committee of the School of Psychology, Cardiff University.

2.3.2 Stimuli

The stimulus set contained angry, happy, and neutral faces (15 male and female faces per condition), as well as 15 scrambled control stimuli. The face images were selected from the NimStim database (Tottenham et al., 2009), which includes both closed-mouth (low arousal) and open-mouth (high arousal) versions of each emotional expression; for this study, we selected closed-mouth neutral expressions, open-mouth happy expressions, and a balanced set of closed-mouth and open-mouth angry expressions, which accounted for the higher arousal associated with angry faces. In practice, this stimulus selection enhances visual differences (i.e. in terms of visible teeth) between the happy and neutral face sets.

The scrambled stimuli were noise images created by combining the average Fourier amplitudes across stimuli with phase information from white noise images of equal size (Perry and Singh, 2014).

All images were 506 x 560 pixels in size and were converted to grayscale (Figure 2.1). To ensure that global low-level properties were matched between stimuli,

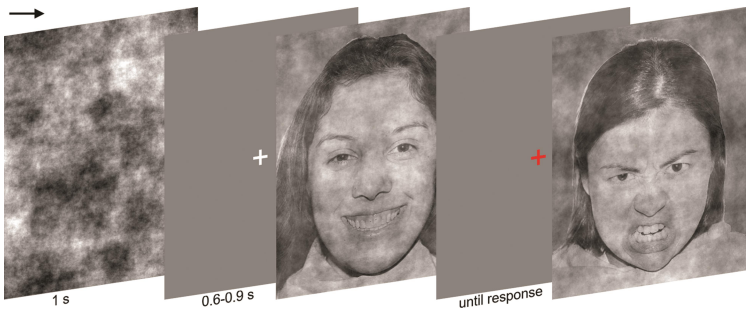

FIGURE 2.1: Experimental paradigm, with examples of one scrambled image and two face stimuli from the NimStim database, after normalization of Fourier amplitudes.

the 2D Fourier amplitude spectrum of each image was set to the average across all stimuli. This was done by calculating the average amplitude spectrum across images in the Fourier domain, and replacing individual amplitude spectra with the average when performing the inverse transformation of each image.

2.3.3 Data Acquisition

All participants underwent a whole-head T1-weighted MRI scan on a General Electric 3 T MRI scanner using a 3D Fast Spoiled Gradient-Recalled-Echo (FSPGR) pulse sequence in an oblique-axial orientation with 1 mm isotropic voxel resolution and a field of view of 256 x 192 x 176 mm.

Whole-head MEG recordings were made using a 275-channel CTF axial gradiometer system at a sampling rate of 600 Hz. Three of the sensors were turned off due to excessive sensor noise and an additional 29 reference channels were recorded for noise rejection purposes. The data were collected in 2.5 s epochs centred around the stimulus onset. A continuous bipolar electrooculogram (EOG) was recorded to aid in offline artefact rejection.

Stimuli were centrally presented on a gamma-corrected Mitsubishi Diamond Pro 2070 CRT monitor with a refresh rate of 100 Hz and a screen resolution of 1024 x 768 pixels. Participants viewed the stimuli from a distance of 2.1 m at a visual angle of 8.3°x 6.1°.

Participants underwent two scanning sessions with up to 5 minutes of break in between. Each session comprised 360 trials, with the 15 images corresponding to each condition presented six times in random order. On each trial, the stimulus

was presented on a mean grey background for 1 s, followed by an interstimulus interval with a duration selected at random from a uniform distribution between 600 and 900 ms (Figure 2.1). A white fixation cross was presented at the centre of the screen throughout the experiment. Participants performed a change detection task to ensure maintained attention: the fixation cross turned red at the start of a pseudorandom 10% of trials (during the inter-stimulus interval) and participants had to press a button using their right index finger in order to continue. The paradigm was implemented using Matlab (The Mathworks, Natick, MA, USA) and the Psychophysics Toolbox (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997).

Participants were seated upright while viewing the stimuli and electromagnetic coils were attached to the nasion and pre-auricular points on the scalp to determine head location. High-resolution digital photographs were used to verify the locations of the fiducial coils and co-register them with the participants' structural MRI scans. Head position was monitored continuously and head motion did not exceed 6.6 mm in any given session.

2.3.4 Data Analysis

Pre-processing

Prior to sensor-space analyses, the data were pre-processed using Matlab and the FieldTrip toolbox (Oostenveld et al., 2011). Trials containing eye movement or muscle artefacts were rejected after visual inspection. One participant was excluded due to excessive artefacts and analysis was performed on the remaining 14 subjects. Across the remaining subjects, the percentage of trials excluded did not exceed 12.7% (mean 40 trials excluded across both sessions, SD 24.3), and the number of trials excluded did not significantly differ between conditions ($P = 0.86, F(2.2, 28.9) = 0.18$).

To monitor head motion, the position of the three fiducial coils relative to a fixed coordinate system on the dewar was continuously recorded during data acquisition. Head motion was quantified as the maximum displacement (difference in position between sample points) of the three coils during any given trial. Using this metric, we excluded trials with maximum motion of any individual coil in

excess of 5 mm. To account for changes in head position, head coil position was changed to the average position across trials for each dataset.

For sensor-space analyses, a 50 Hz comb filter was used to remove the mains noise and its harmonics and baseline correction was applied using a time window of 500 ms prior to stimulus onset.

Event-related field (ERF) analysis

We inspected event-related fields in order to examine differences between conditions present in single-channel responses. The data were bandpass-filtered between 0.5 and 30 Hz using fourth-order IIR Butterworth filters. ERFs were realigned to a common sensor position (Knösche, 2002) and averaged across subjects. We then identified three time windows of interest based on local minima in the global field power across all face conditions (Figure 2.3D; Perry and Singh, 2014): ~60-127 ms (M100), 127-173 ms (M170), and 173-317 ms (M220). ERF responses were averaged within each time window of interest. For each time window, we tested for differences between trial-averaged responses to neutral and scrambled faces and between emotional faces using a paired t-test and a repeated-measures ANOVA respectively and randomization testing (5000 iterations, corrected using the maximal statistic distribution across sensors).

MVPA pre-processing and feature selection

Sensor space: Prior to sensor-space MVPA analyses, the data were averaged in groups of 5 trials to improve SNR (Grootswagers et al., 2017; Isik et al., 2014). The number of observations was not significantly different between conditions (Angry: 33.6 ± 1.6; Happy: 33.4 ± 1.4; Neutral: 33.5 ± 1.1; Scrambled: 33.6 ± 1; $F(3, 13) = 0.64, P = 0.59$). To assess differences between responses to neutral and emotional faces as well as between different emotional expressions, binary classification was applied to all pairs of emotional conditions.

We assessed the presence, latency and coarse spatial location of expression-specific information at the sensor level by performing within-subject time-resolved classification on data from four anatomically defined sensor sets (occipital, temporal, parietal and frontocentral; Figure 2.6). MVPA was performed at each sampled

time point (every ~1.67 ms) between 0.5 s pre-stimulus onset and 1 s post-stimulus onset. Compared to a whole-brain approach, this method served to reduce the number of features while also providing some spatial information.

To maximize the number of informative features used as input to the classifier, we conducted an additional sensor-space MVPA analysis in which feature selection was performed based on differences between faces and scrambled stimuli. This ensured unbiased feature selection based on an orthogonal contrast and led to the selection of sensors responding most strongly to faces, in order to maximize the interpretability of our results.

To determine sensors responding differentially to faces and scrambled stimuli we used a searchlight MVPA approach (Tsuchiya et al., 2008), whereby each MEG channel and its neighbouring sensors, defined according to a Fieldtrip template based on the CTF 275-sensor array configuration, were entered separately into the MVPA analysis. Searchlights were defined to include only sensors directly connected to the centroid according to the template, and searchlight size thus ranged between 4 and 10 sensors (mean 7.36, SD 1.12). The analysis was performed using time windows of approximately 16 ms (10 sampled time points) and stratified five-fold cross-validation was used to evaluate classification performance. Data from the cluster centroids found to achieve above-chance decoding performance in 100% of participants (regardless of latency) were then entered into the three emotional expression classification analyses (Figure 2.6B).

To ensure we captured informative sensors in the expression decoding analysis, two additional feature selection methods based on the face vs scrambled contrast were performed, yielding: (1) 15 sensors found to exhibit significant differences in ERFs between faces and scrambled stimuli in any of the three time windows tested; and (2) a combined set of 55 sensors identified through the MVPA and ERF-based feature selection methods.

Source space: To move beyond the limitations of sensor-space spatial inference in our MVPA analysis and alleviate concerns of signal leakage, head motion and inter-individual variability (Zhang et al., 2016b), the data were projected into source

space using the linearly constrained minimum variance (LCMV) beamformer (Hillebrand et al., 2005; Van Veen et al., 1997). Beamformer weights were normalized by their vector norm to alleviate the depth bias of MEG source reconstruction (Hillebrand et al., 2012). The participant's MRI was used to define the source space with an isotropic resolution of 6 mm and the output for each location was independently derived as a weighted sum of all MEG sensor signals using the optimal source orientation (Sekihara et al., 2004).

The data were projected into source space using trials from all conditions filtered between 0.1 and 100 Hz to calculate the beamformer weights. A frequency analysis was performed using the multitaper method based on Hanning tapers in order to identify the peak virtual channel in each of 84 Automated Anatomical Labeling (AAL; Tzourio-Mazoyer et al., 2002) atlas-based ROIs (excluding the cerebellum and some deep structures; see Figure 2.7A). The classifier input consisted of the raw time-series for each of the 84 virtual sensors, baseline corrected and averaged in groups of 5 trials to improve SNR. Decoding was performed per sampled time point as in sensor space.

To assess whether the MVPA effect found at the source level was also present in univariate responses when eliminating the issues associated with sensor-level analyses, we also calculated evoked responses (trial averages) for the peak sources in each of the 84 ROIs used in the MVPA source-space analysis, filtered between 0.5 and 30 Hz. These were subjected to statistical analysis using the time windows identified at sensor level (2.3.4).

Classifier training and testing

A linear L1 soft-margin Support Vector Machine (SVM) classifier was implemented in Matlab using the Machine Learning and Statistics Toolbox and the Bioinformatics Toolbox (Mathworks, Inc.). Stratified five-fold cross-validation was implemented for training and testing and data points were standardized using the mean and standard deviation of the training set. The box constraint parameter c, which controls the maximum penalty imposed on margin-violating observations, was set to 1.

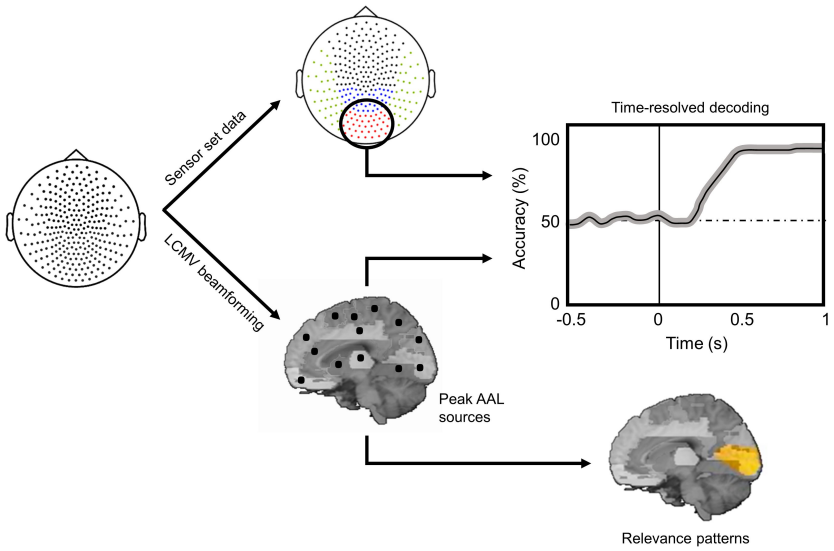

FIGURE 2.2: MPA analysis framework used in this chapter. Time-resolved decoding was performed on (1) sensor-level data from a selected subset (anatomical or data-driven), and (2) source-space data from peak broadband sources in 84 AAL regions.

Computing relevance patterns in source space

For each decoding problem, participant and time point, the SVM model based on source-space data was retrained on the full dataset to obtain the final model and calculate the weight vector. The weight vector for a linear SVM is based on the Lagrange multipliers assigned to each data point (Chapter 1). To achieve interpretable spatial patterns (Haynes, 2015), feature weights were transformed into relevance patterns through multiplication by the data covariance matrix (Haufe et al., 2014). This allowed us to dynamically and directly assess the relative importance of all virtual electrodes used in source-space decoding, as each ROI was represented by one feature and each decoding iteration was run on the whole brain.

Significance testing

To quantify classifier performance, we report average accuracies across subjects (proportions of correctly classified cases), as well as F1 scores (harmonic means of precision and sensitivity) and bias-corrected and accelerated bootstrap confidence intervals using 1000 resampling iterations (Efron and Tibshirani, 1986; Efron, 1987).

Significance was assessed using randomization testing. For each individual dataset, labels were shuffled 1000 times across the training and test sets to create

an empirical null distribution and classification was performed on the randomized data at the time point achieving the highest classification performance across subjects on the real data. For searchlight classification, p-values were calculated for each subject and combined to achieve a group map quantifying the proportion of subjects achieving significance in each searchlight (Pereira and Botvinick, 2011). For all other analyses, randomization was performed within-subject and empirical null distributions were calculated in an identical manner as the observed statistic (i.e. average accuracy over subjects).

To correct for multiple comparisons, we tested average accuracies against the omnibus null hypothesis by thresholding using the maximum accuracy distribution (Nichols and Holmes, 2001; Singh et al., 2003). For classification on different sensor sets, this was done by selecting the maximum average performance across sensor sets to create a null empirical distribution. For searchlight classification, p-values were thresholded using the maximum performance across sensor clusters. For sensor-space classification based on feature selection and for source-space classification, p-values were adjusted using the false discovery rate and cluster-corrected across time. Permutation p-values were calculated taking the observed statistic into account, using the conservative estimate $p = (b+1)/(m+1)$, where b is the number of simulated statistics greater than or equal to the observed statistic and m is the number of simulations (Phipson and Smyth, 2010).

To identify the ROIs significantly contributing to decoding performance in source space, permutation testing (5000 sign-flipping iterations) was applied to baselined mean relevance patterns for each ROI and time window. P-values were corrected for multiple comparisons using the maximum statistic distribution across ROIs, and a further Bonferroni correction was applied to account for the multiple time windows tested.

Control analyses

Decoding was also performed on the EOG timeseries to control for the possibility of eye movements driving decoding performance, and the impact of low-level features was assessed by applying classifiers to image properties, specifically pixel intensity levels and the spatial envelope obtained using the GIST descriptor (Oliva

and Torralba, 2001). The latter consisted of 256 values for each image, obtained by applying Gabor filters at different orientations and positions to extract the average orientation energy. Although it was originally designed to capture scene properties and is perhaps less suited to extracting face information, the spatial envelope is a holistic representation of low- and mid-level properties; it thus summarizes the orientation information in our stimuli without extracting face-specific features that would be expected to encode emotion and determine expression recognition.

2.4 Results

2.4.1 Evoked responses to faces

When assessing the effect of emotional expression on event-related fields (Figure 2.3A-D), we found no modulation of any of the three ERF components ($F(2,26) < 9.37$, $P > 0.061$ across all three comparisons). Conversely, we found significant differences between responses to faces and scrambled faces at the M170 latency ($t(13) > 5.43$, $P < 0.0078$; maximum $t(13) = 7.17$, $P = 0.0008$) and at the M220 latency ($t(13) > 5.38$, $P < 0.0099$; maximum $t(13) = 6.54$, $P = 0.0016$). At the M100 latency, no differences survived correction for multiple comparisons ($t(13) < 4.41$, $P > 0.04$).

Univariate responses at the source level showed a similar pattern (Figure 2.3E-F). Statistical analysis of the ROI-averaged response revealed a significant difference between faces and scrambled stimuli only in the M170 window ($P = 0.0012$, $t(13) = 4.89$; paired T-test and randomization testing using 5,000 iterations). Tests performed at each ROI were inconclusive ($P > 0.09$, $t(13) < 4.1$). We note here that the selection of one source per ROI and the number of comparisons performed are likely to be the cause of these results. Furthermore, to assess differences in expression, we performed repeated-measures ANOVAs and randomization testing (5,000 iterations) on both ROI-averaged data and at each ROI separately using the same three time windows of interest. Neither of these approaches revealed significant results ($P > 0.22$, $F(2,26) < 1.58$, and $P > 0.25$, $F(2,26) < 6.5$ respectively).

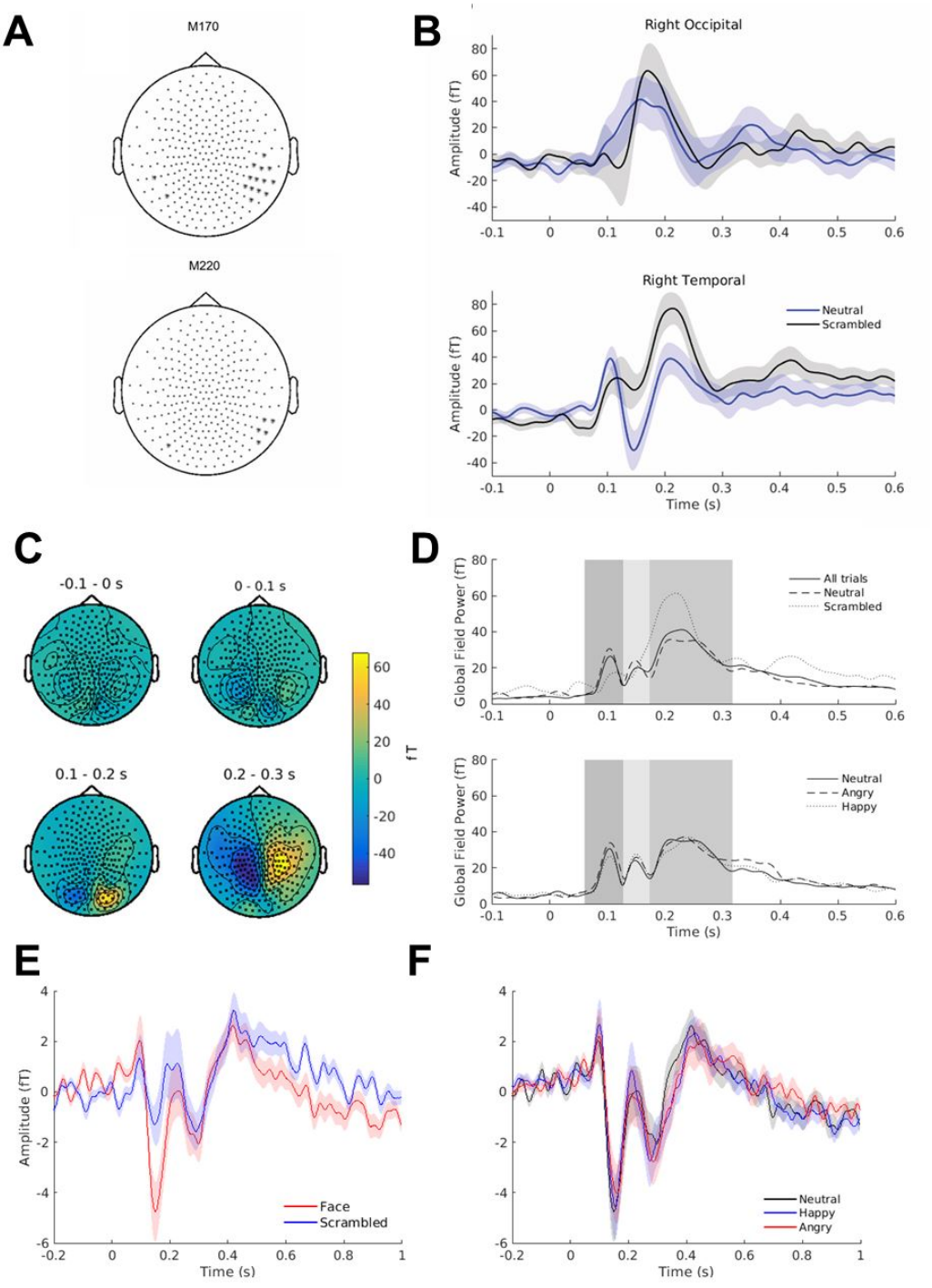

FIGURE 2.3: Evoked responses to faces and expressions. **A**. Sensors exhibiting significant differences to faces compared to scrambled stimuli (marked with asterisks) at the M170 and M220 latencies (P<0.01). **B**. Timecourses of the evoked responses to neutral faces and scrambled stimuli from right occipital and temporal sensors averaged across subjects ($\pm SEM$). **C**. Topographical distribution of the grand average ERF amplitudes from all axial gradiometers across the three face conditions. **D**. Global field power of the grand average ERF across all trials and for each condition. Shaded areas show windows of interest in the ERF analysis. **E-F**. Grand average evoked responses ($\pm SEM$) over 84 ROIs and all 14 subjects in source space.

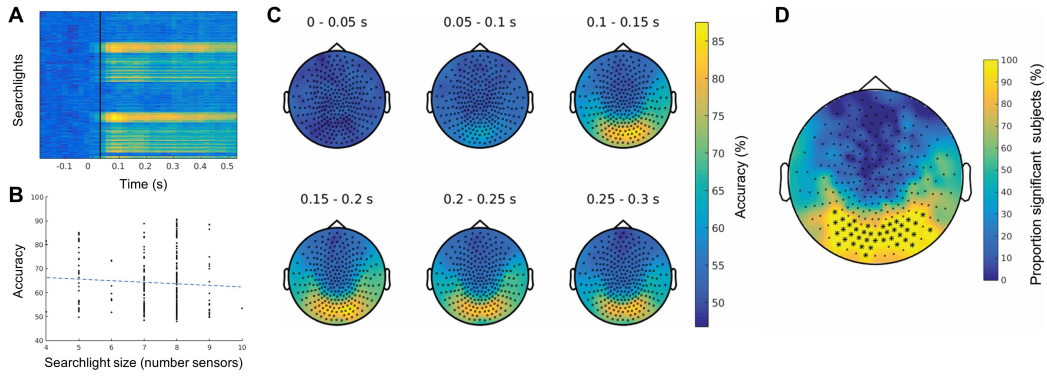

FIGURE 2.4: Sensor-space decoding of faces vs scrambled stimuli. **A**. Time-resolved decoding accuracy for all searchlights. The black vertical line marks the onset of above-chance decoding (80-110 ms). **B**. Scatterplot of averaged accuracies across subjects (133-150 ms) for all searchlight sizes, showing no relationship between searchlight size and accuracy. **C**. As in **A**, but plotted on the MEG sensor layout and averaged over 50 ms time windows. **D**. Proportion of participants achieving above-chance decoding at each searchlight regardless of latency. Sensors significant in all subjects and selected for further analysis are marked with asterisks.

2.4.2 MVPA results: decoding faces and scrambled stimuli

A searchlight MVPA analysis was performed on the face vs scrambled decoding problem to identify sensors of interest for emotional expression classification. Faces were decoded above chance starting at ~80 ms at occipito-temporal sensors (Figure 2.4A). We thus identified a set of 40 occipito-temporal sensors achieving above-chance decoding performance in all participants at any time point after stimulus onset (Figure 2.4C). Note that although searchlights included neighbouring sensors around a centroid and thus varied in size, there was no correlation between searchlight size and decoding accuracy (Pearson's $\rho = -0.059$, $P = 0.33$; Figure 2.4B).

Source-space face decoding showed a similarly early onset (~100 ms), with slightly lower decoding accuracies. Relevance patterns based on classifier weights highlighted the visual cortex and fusiform gyrus between 100-200 ms post-stimulus onset (coinciding with the M170 effects found in the ERF analysis; Figure 2.5).

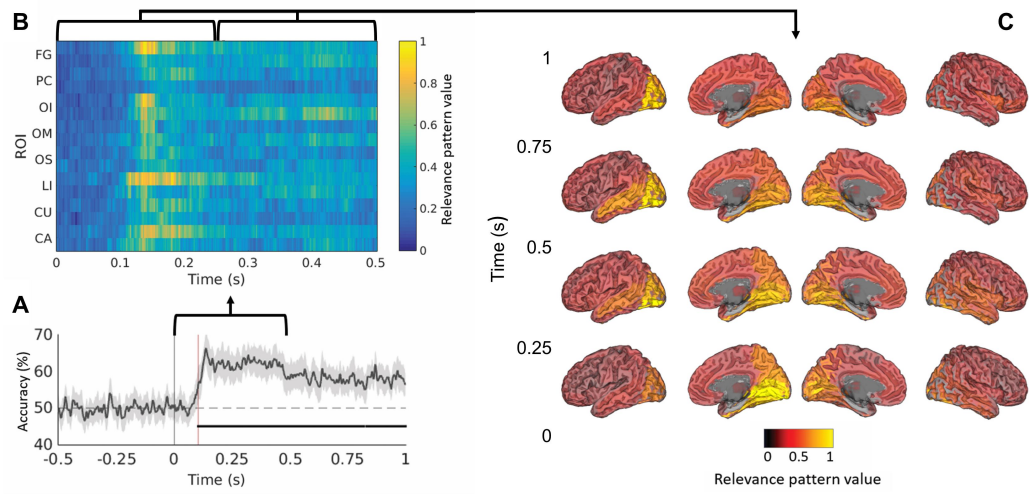

FIGURE 2.5: Source-space decoding of faces vs scrambled stimuli. **A**. Decoding accuracy for the face vs scrambled problem in source space with 95% CI and significant decoding time window (black horizontal line, starting at 100 ms). **B**. Patterns derived from broadband source-space decoding of faces and scrambled stimuli for 8 key ROIs for the 0–500 ms time window after stimulus onset. **C**. Whole-brain patterns averaged across 250 ms windows and plotted on the semi-inflated MNI template brain. Bilateral ROI labels: CA: calcarine cortex; CU: cuneus; LI: lingual gyrus; OS: occipital superior; OM: occipital medial; OI: occipital inferior; PC: precuneus; FG: fusiform gyrus.

2.4.3 MVPA results: decoding emotional faces

Sensor space decoding

When using anatomically defined sensor sets to define the feature space, MEG data from occipital sensors successfully discriminated angry and neutral faces (at 93 ms post-stimulus onset), as well as angry and happy faces (at 113 ms post-stimulus onset). The classification of happy and neutral faces was delayed and showed a weaker effect, which reached significance for a brief time window at 278 ms. The temporal sensor set successfully decoded angry vs neutral faces starting at 262 ms. Other sensor sets did not achieve successful classification (Figure 2.6A). The maximum average accuracy across subjects was achieved in the occipital sensor set decoding of angry vs neutral faces (65.39%, bootstrap 95% CI [60.83%, 69.51%]; Table 2.1).

Feature selection of sensors that successfully decoded faces vs scrambled stimuli marginally improved classification performance (Table 2.1) and led to above-chance accuracy on all three binary classification problems, starting at ~100 ms for

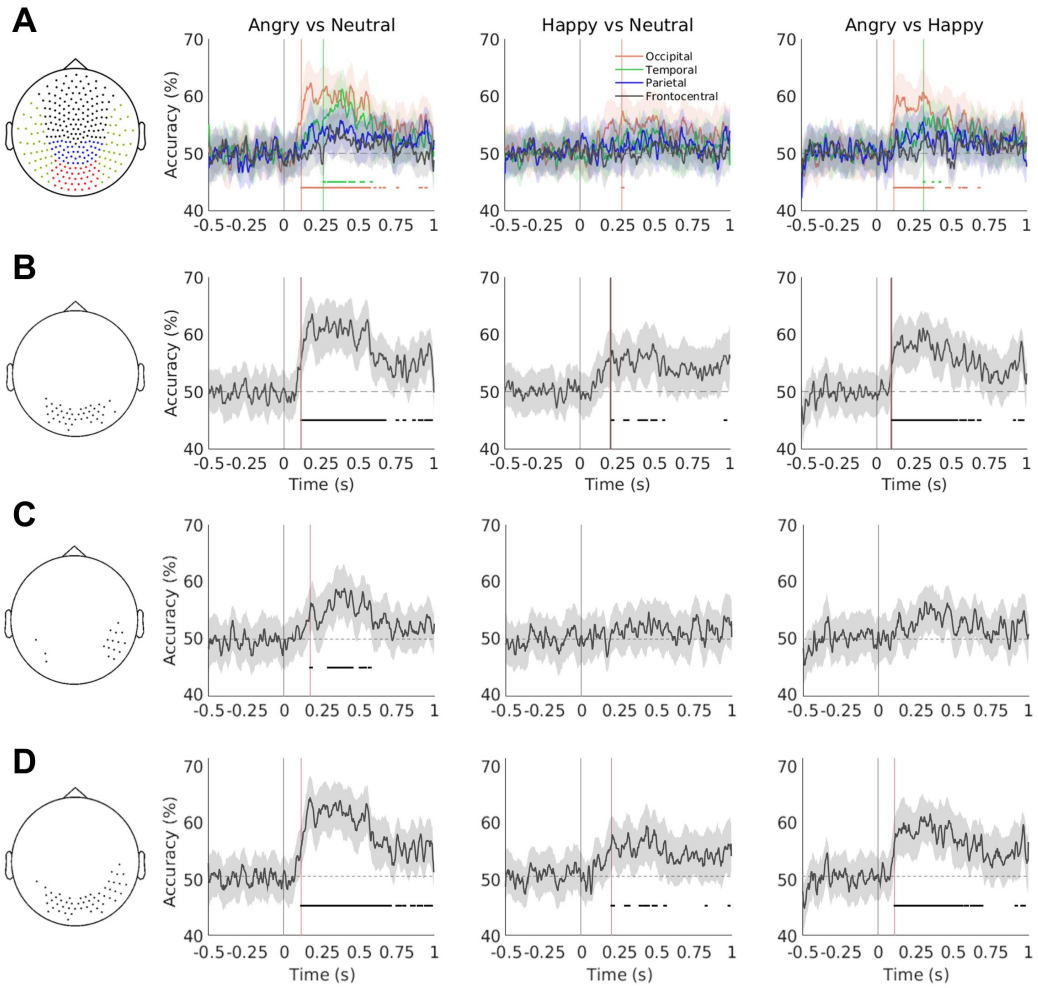

FIGURE 2.6: **A**. Accuracy traces averaged across participants for each emotion classification problem and each of the four sensor sets (shown in the left-hand plot). The vertical lines mark the stimulus onset and the shaded areas depict 95% bootstrapped CIs. The horizontal lines represent clusters of at least five significant timepoints (FDR-corrected P<0.05). Significant decoding onset is marked with vertical lines (at 100 ms for the angry vs. neutral/happy face decoding using occipital sensors). Accuracy traces were smoothed with a 10-point moving average for visualization only. The remaining panels show time-resolved accuracies using: **B**. the sensor set based on the searchlight feature selection method (shown in the left-hand plot); **C**. the ERF-based sensor set; **D**. the joint sensor set (based on both MVPA and ERF results). Different methods of feature selection lead to similar results.

.

angry faces and at ~200 ms for happy and neutral faces (Figure 2.4B). Using the sensors exhibiting an ERF response to faces delayed the decoding onset to 175 ms (maximum accuracy 61.68%, CI [58.68%, 64.93%]), highlighting the difference in information content between evoked responses and multivariate patterns. Finally, using the joint sensor set achieved similar results to occipital sensor decoding (decoding onset at 116 ms, maximum accuracy 65.59%, CI [60.97%, 69.5%]).

Source space decoding

We used 84 peak virtual electrodes in AAL atlas-based ROIs to perform whole-brain decoding of emotional expression in source space. Angry faces were decodable from neutral faces at 155 ms and from happy faces at 300 ms, while happy and neutral faces were less successfully decoded, with a non-significant peak at 363 ms.

Later onsets of significant effects in source space are likely to be due to the whole-brain approach and the subsequently lower accuracies obtained in source space. Accuracy may have been decreased by the higher number of features and by our choice of one peak timecourse per ROI as input to the classification, which may have filtered out informative signal. However, as optimizing accuracy was not the main goal of this study, our method offers interpretability advantages, such as the ability to assess the relative roles of different ROIs without the confound of unequal ROI or feature vector sizes. Although feature selection could improve classification performance, we decided against optimizing accuracy in favour of deriving whole-brain maps from classifier weights.

Source-space relevance patterns

To assess ROI contributions to source-space decoding performance, classifier weights were converted into relevance patterns and then averaged across subjects and over time using 100 ms time windows. Relevance patterns attributed a key role to occipital regions within 200 ms of stimulus onset, with temporal and frontal regions contributing information at later stages (Figure 2.7). This was confirmed by permutation testing results, which highlighted the role of the right lingual gyrus in discriminating angry and neutral faces within 200 ms (Figure 2.8). Information in the left calcarine sulcus and inferior occipital gyrus (with a potential source in the

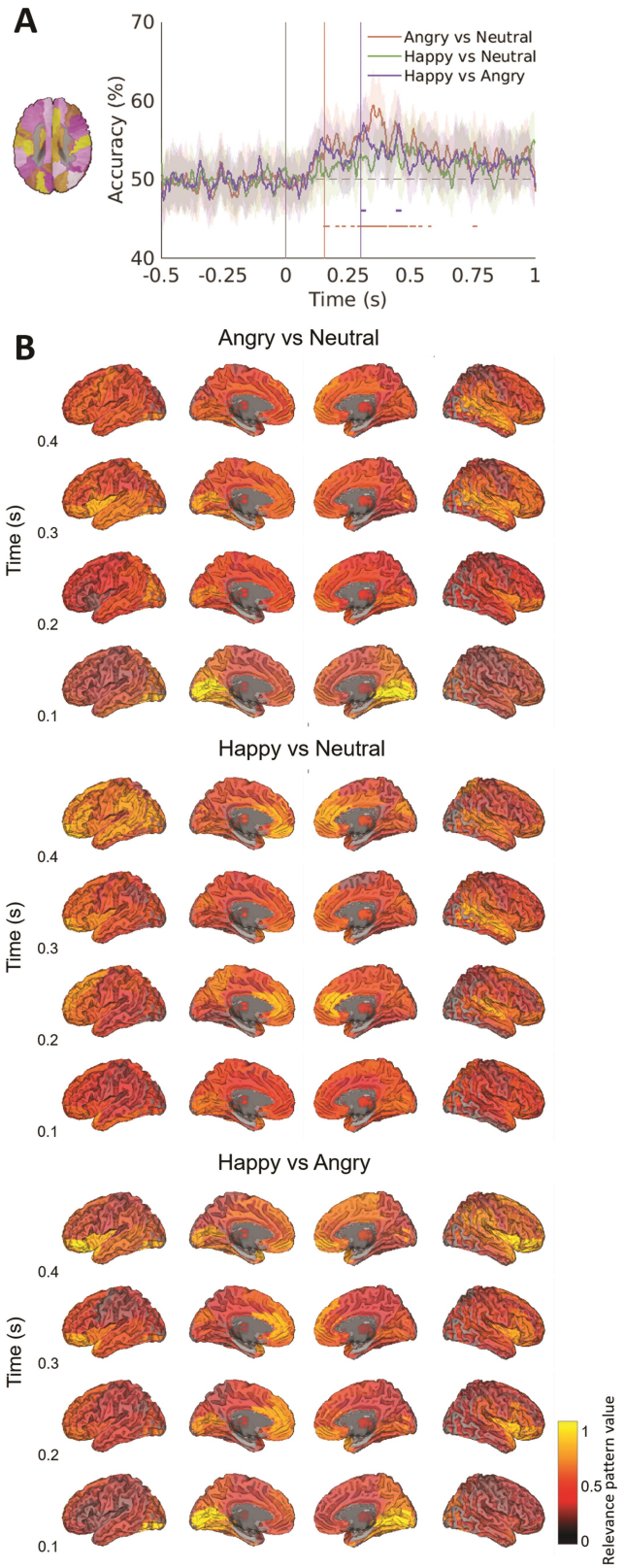

FIGURE 2.7: **A**. Accuracy traces averaged across participants for each emotion classification problem in source space using the 84 AAL atlas-based ROIs (shown in the left-hand plot). **B**. Broadband relevance patterns derived from classifier weights in source space for all three decoding problems, averaged across subjects and 100 ms time windows, baselined and normalized, mapped on the semi-inflated MNI template brain (100-500 ms). Patterns show the relative role of each ROI in decoding without statistical testing.

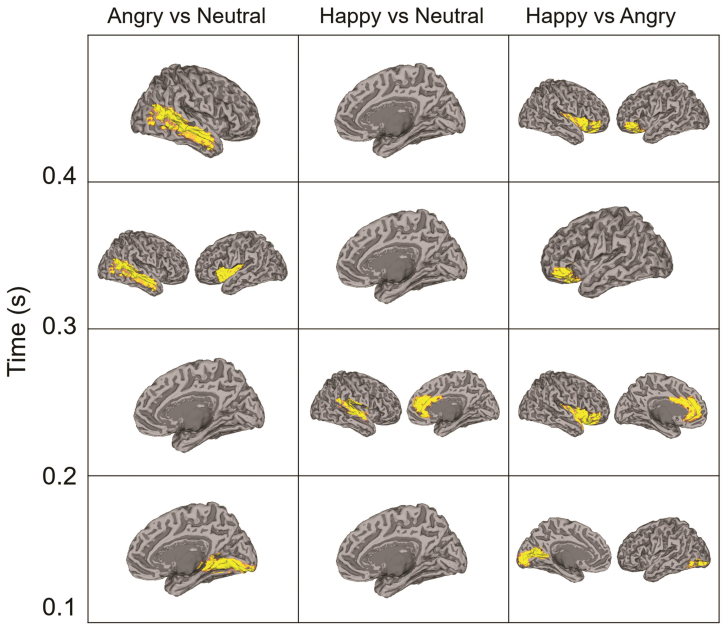

FIGURE 2.8: Results obtained from randomization testing of the relevance patterns shown in Figure 2.7 for each decoding problem and time window between 100 and 500 ms. Highlighted ROIs were assigned significant weights (P<0.05 corrected).

occipital face area) appeared to differentiate angry and happy faces, while areas in the temporal, insular and inferior orbitofrontal cortices were involved at later stages in all three classification problems.

2.4.4 Control analyses

For all three decoding problems, time-resolved decoding performed on the EOG timeseries (using 25 time points from each of the two EOG channels as features) achieved a maximum accuracy no higher than 50.9% (bootstrapped 95% CIs [47.75%, 52.6%]). Classification performed on the entire EOG timeseries did not exceed 52.49% (CI [48.6%, 56.3%]). This suggests that decoding results were unlikely to be driven by eye movement artefacts.

Binary classification between conditions based on raw image properties (intensity levels per pixel ranging between 0 and 1, mean 0.53, SD 0.16) was not significantly above chance, although suggestive for one decoding problem (happy versus neutral: 33% accuracy, P=0.9; angry versus neutral: 60% accuracy, P=0.24; and angry versus happy: 70% accuracy, P=0.053, randomization testing).

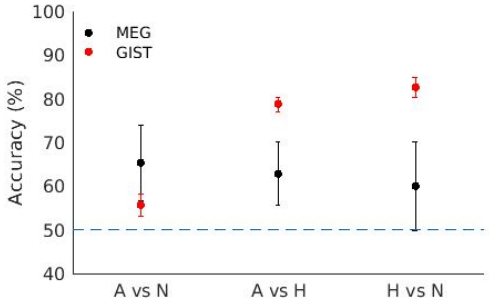

FIGURE 2.9: Decoding results obtained using: (1) the MEG occipital sensor set at peak time point across subjects (MEG); (2) the spatial envelope calculated using the GIST descriptor (GIST). Error bars represent bootstrap 95% CIs based on classification across subjects/cross-validation iterations. The blue dashed line marks the theoretical chance level (although note that the angry vs neutral GIST-based classification does not exceed the empirically estimated chance level). A: angry; H: happy; N: neutral.

Finally, we performed binary classification between pairs of emotional expression conditions, using the spatial envelope values calculated using the GIST descriptor for each image. Two of the decoding problems were successfully solved (happy versus neutral: 82.6% accuracy, P=0.0032, happy versus angry: 78.7%, accuracy, P=0.0062), while angry faces could not be decoded from neutral faces (55.67% accuracy, P=0.33). This suggests that in our stimulus set, visual properties distinguish happy faces from neutral and angry faces (unsurprisingly, given the consistency in happy expressions), while angry faces are not easily distinguishable from neutral faces. These results stand in contrast to results from MEG decoding (Figure 2.9), which follow an inverse pattern, with the highest accuracies obtained when decoding angry and neutral faces.

Despite our efforts to match Fourier amplitudes between stimuli, low-level differences between expressions remain that may contribute to the results and that can be expected to play an important role in expression recognition. However, the control analyses suggest that our MEG results cannot be readily explained by global differences in spatial envelope or pixel intensities. The increase in accuracy when decoding angry faces from other expressions (~100 ms), while likely to be based on low-level information associated to emotional expression, is not easily explained by unrelated visual properties.

TABLE 2.1: Expression decoding results in sensor and source space

Angry vs Neutral	Occipital	Temporal	Parietal	Frontocentral	Selected	Source space
Max accuracy	65.39%	63.68%	58.62%	57.52%	65.96%	61.13%
95% CI	60.83%, 69.51%	58.91%, 68.68%	56.20%, 61.59%	52.28%, 60.23%	62.03%, 69.11%	57.41%, 64.77%
Max F1 score	0.653	0.636	0.585	0.573	0.659	0.611
Peak time point	267 ms	388 ms	947 ms	618 ms	185 ms	376 ms
Decoding onset	93 ms	262 ms	N/A	N/A	113 ms	155 ms
Happy vs Neutral						
Max accuracy	59.97%	58.23%	58.14%	57.30%	60.65%	58.98%
95% CI	55.11%, 65.27%	55.37%, 61.01%	54.36%, 63.28%	53.32%, 61.56%	57.05%, 65.22%	56.31%, 61.12%
Max F1 score	0.599	0.581	0.58	0.572	0.605	0.589
Peak time point	485 ms	315 ms	673 ms	637 ms	481 ms	363 ms
Decoding onset	278 ms	N/A	N/A	N/A	205 ms	N/A
Happy vs Angry						
Max accuracy	62.83%	62.29%	57.97%	57.02%	64.03%	60.93%
95% CI	59.70%, 66.88%	57.18%, 65.60%	54.43%, 63.06%	53.21%, 60.55%	59.32%, 69.87%	57.29%, 64.91%
Max F1 score	0.628	0.621	0.578	0.568	0.639	0.609
Peak time point	332 ms	468 ms	465 ms	403 ms	313 ms	455ms
Decoding onset	113 ms	N/A	N/A	N/A	98 ms	301 ms

2.5 Discussion

In this chapter, we used sensor-space and source-localized MEG data and data-driven multivariate methods to explore the spatiotemporal dynamics of emotional face processing. We report three main findings based on our analyses. First, the emotional valence of faces (especially angry expressions) can be robustly decoded based on data from occipito-temporal sensors, as well as whole-brain source-space data. Second, information related to emotional face category is available as early as 90 ms post-stimulus onset, despite a lack of effects in trial-averaged ERFs. Third, data-driven relevance maps link different stages in expression perception to visual cortex areas (early stages) and higher-level temporal and frontal cortices (later stages).

2.5.1 Early processing of facial expressions

Although we found no modulation of trial-averaged ERF components by emotional expression, our ERF analysis revealed a face response over temporal sensors at the M170 and M220 latencies and no face-specific M100 component, in line with previous studies using matched control stimuli and similar designs (Perry and Singh,

2014; Rossion and Caharel, 2011). On the other hand, an early occipito-temporal response to faces at M100 latencies was revealed in the MVPA analysis. Together, these results appear to point to different components in face processing – an early occipital effect not present in the trial-averaged ERFs, and a later, mainly right-lateralized temporal effect. Note that although the sensors contributing the most information to the MVPA analysis are different to the sensors identified in ERF analysis, the latter set of sensors do perform above chance when used in MVPA analysis in a majority of subjects (Figure 2.6C); the increased heterogeneity can be explained by lower cross-subject consistency at the sensor level of a late, higher-level response.

Using MVPA, we were able to identify expression-related information at early latencies in the sensor-level MEG data. Expression (angry and neutral/happy faces) could be decoded at 93 ms and 113 ms respectively, only 10-30 ms later than faces were decoded from scrambled stimuli, and earlier than latencies reported by previous ERP studies (even by those showing emotional modulation of P1; e.g. Aguado et al., 2012). Such early latencies are consistent with neurophysiological investigations in primates: for example, multivariate analysis of local field potential (LFP) data in monkeys has shown early categorisation of faces at 60-90 ms (Cauchoix et al., 2012), while face-selective cells in primate temporal cortex respond to faces or facial features at 80-100 ms (Hasselmo et al., 1989; Perrett et al., 1982). Our results add to recent evidence of rapid visual categorization occurring during the early stages of ventral stream visual processing (Cauchoix et al., 2016; Clarke et al., 2013) and suggest that this extends beyond low-level properties. Moreover, we reveal differences in patterns that can be detected in the absence of trial-averaged ERF effects. Such differences, together with method heterogeneity, could explain previous mixed results in ERF studies, and speak to the sensitivity advantage of MVPA. In light of this, similar MVPA approaches will be used in the next two chapters of this thesis to answer more specific questions about the computations underpinning expression processing.

On a different note, the lower performance and later onset of happy versus neutral face decoding suggests a categorization advantage inherent in angry expressions. Angry faces were decoded from both happy and neutral faces almost

simultaneously, suggesting a bias related to threat and not to emotion in general. This points to a system preferentially responsive to threat, consistent with models placing conflict resolution at the core of social interaction (Waal, 2000). Furthermore, the whole-brain, data-driven analysis pipeline employed here revealed this bias without entailing assumptions about the temporal or spatial location of an effect.

2.5.2 Spatial patterns of expression-related information

We implemented an atlas-based approach to source-space decoding in order to improve the interpretability of the resulting maps and to facilitate cross-modality comparisons (Hillebrand et al., 2012). This approach has been successfully applied to resting-state MEG studies (e.g. Brookes et al., 2016) and, together with the selection of a peak source per ROI, allowed us to increase the computation speed of our whole-brain decoding analysis, while at the same time reducing data dimensionality and allowing for direct comparison between ROIs. The relevance patterns in this study were stronger at time points corresponding to accuracy increases (starting at ~100 ms), but we refrain from directly linking the two because we did not optimize accuracy in this study.

When decoding angry and neutral/happy faces, early differential processing was localized to the calcarine, lingual and inferior occipital ROIs, starting at approximately 100 ms post-stimulus onset (Figure 2.7). Other occipital ROIs showed a weaker contribution to decoding, with patterns later spanning a range of temporal and frontal areas. Early patterns differentiating neutral and happy faces were weaker (as confirmed by the lack of significant ROIs for this problem in the first 200 ms, and explained by the low decoding accuracy), but evolved similarly over time (Figure 2.8). Strong patterns in the early visual cortex and the occipital face area may be evidence of preferential threat processing based on coarse visual cues which are rapidly decoded and forwarded to higher-level regions. Emotional modulation in the visual cortex has previously been reported (Fusar-Poli et al., 2009; Herrmann et al., 2008; Padmala and Pessoa, 2008), and the current results suggest that this effect occurs within 200 ms of face onset.

The traditional model postulating different pathways for processing static facial features (such as identity) and changeable features (such as expression; Bruce and Young, 1986; Haxby et al., 2000) has been challenged by mounting evidence of interaction between the two systems (Rivolta et al., 2016). Despite their coarse spatial resolution, our results suggest that face-responsive areas, including those thought to process identity, respond to emotional expression. The OFA/inferior occipital gyrus appears to be involved at an early stage, while the fusiform gyrus and the superior temporal ROIs (locations of the FFA and STS) are recruited at later time points. These results are in line with previous fMRI MVPA studies demonstrating above-chance expression decoding in all face-selective regions (Wegrzyn et al., 2015) and particularly in the FFA, STS and amygdala, in the absence of univariate effects (Zhang et al., 2016a). Later time windows are characterized by patterns in the insular, prefrontal and orbitofrontal cortices, previously associated with emotional processing especially at the later stages of integration and evaluation (Chikazoe et al., 2014; Phan et al., 2002).

The timing of expression processing as evaluated with MEG MVPA can offer indirect evidence of the hierarchy of the modules involved. In this chapter, the short latencies of emotional face discrimination in visual cortex can be interpreted as supporting a feedforward model of expression processing (Lohse et al., 2016; Wang et al., 2016). Since we find the earliest differential effects in early visual cortex (within 100 ms), this appears to be somewhat inconsistent with the preferential relaying of expression information via the subcortical route to the amygdala (Pessoa and Adolphs, 2011), although subcortical structures were not directly investigated here. However, the current data are not incompatible with the possibility of a subcortical route with no preference to expression (Garvert et al., 2014; McFadyen et al., 2017).

2.5.3 What does successful emotional face decoding tell us?

Naturalistic and high-level stimuli, although appropriate for linking perception to cognitive processing, may give rise to ambiguities in interpretation. In this experiment, Fourier amplitudes were matched across stimuli to the detriment of their

naturalistic qualities. As emotional processing can encompass several distinct processes, a passive viewing paradigm was employed to eliminate task-related or top-down attention effects. Attentional effects would thus be expected to arise due to emotional salience in a bottom-up fashion compatible with our results.

The matching of some low-level properties does not preclude the existence of local differences between images that are likely to play a part in early decoding. However, the fact that angry faces are decoded more successfully than happy/neutral faces points to their relevance rather than to non-specific decoding based on low-level properties; for example, happy faces could be expected to be successfully decoded by a low-level classifier due to their consistent smiles, as suggested by their successful decoding based on spatial envelope features. Furthermore, successful classification based on sensors that discriminate between faces and scrambled stimuli adds to the evidence that our data do reflect face processing. It is likely that local low-level properties play a part in decoding (especially in early time windows and low-level visual areas); however, such properties can be viewed as informative in the emergence of high-level categories. Thus, these results suggest that behaviourally relevant (threat-related) low-level cues are detected and relayed preferentially compared to benign emotional cues.

One limitation of this experiment is the fact that cross-exemplar decoding could not be performed in order to assess classifier generalization to a novel set of stimuli, as the occurrence of each exemplar was not recorded in this paradigm. Thus, there is a concern about the classifier potentially exploiting stimulus repetitions in order to successfully classify the two categories. However, as repetition numbers were balanced across conditions, we would expect this concern to affect all three decoding problems equally. As the control analyses do not point to the angry faces as more classifiable in terms of low-level properties, the successful decoding of angry faces from MEG data is consistent with their behavioural relevance and not with recognition of individual exemplars and stimulus properties. In subsequent experiments described in Chapters 3 and 4, this concern was addressed using cross-decoding of emotional expressions from MEG data. In Chapter 3 in particular, we show remarkably similar temporal dynamics (decoding of both face presence and expression at ~100 ms) using a cross-exemplar decoding approach, although the

threat advantage characteristic appears to depend on task context.

Furthermore, the use of stimulus repetitions to achieve robust responses to a limited stimulus set poses the concern of potential differences in repetition suppression effects. Such effects have been shown to covary with a number of factors, including time lag, task type, stimulus familiarity and valence (Morel et al., 2009). In particular, a stronger repetition suppression effect was shown for fearful faces than for neutral faces in both fMRI and MEG (Ishai et al., 2004; Ishai et al., 2006), although this effect was only present for target faces that were the object of a working memory task. On the other hand, repetition suppression was shown to be absent for happy faces and reduced for angry faces as compared with neutral faces in an fMRI study with an implicit paradigm (Suzuki et al., 2011). Such a pattern is inconsistent with a large contribution of repetition suppression effects to the current results. Furthermore, previous studies have shown differential repetition effects in evoked response potentials, while evoked responses in the current data revealed no differences between expressions.

Finally, despite the advantages of the information mapping approach, challenges remain in the interpretation of decoding results (Chapter 1). Although patterns derived from classifier weights indicate the availability of decodable information, it is difficult to assess the type of information used by the classifier or whether this same information is functionally relevant. However, the results are validated by existing models of emotional face processing, whereby large-scale differences in spatial patterns over time may be elicited by different pathways involved in processing neutral and emotional/ threat-related and benign stimuli. On the other hand, the role played by individual ROIs in decoding can be interpreted as reflecting differences in neuronal population activity, as suggested by fMRI, MEG and electrophysiological investigations establishing correlations between face-selective cell activity, the BOLD signal (Hasselmo et al., 1989; Tsao et al., 2006) and gamma oscillations (Muthukumaraswamy and Singh, 2008; Perry, 2016; Perry and Singh, 2014). It is likely that different regions contribute different types of discriminating information and further study is needed to tease apart the underlying neural activity. While the overlap in areas between classification problems and the distributed nature of expression-related information hint at the existence of a core system that

efficiently identifies and relays emotional cues, the spatial resolution of these data is too coarse to make strong claims about the structure of this system.

The findings discussed here extend beyond successful decoding of emotional stimuli to reveal a system optimised for rapid processing of emotional content in faces and particularly tuned to angry expressions. Decoding timecourses and relevance patterns indicate that affective information is rapidly relayed between early visual cortex and higher-level areas involved in evaluation, suggesting that in a passive viewing paradigm, behavioural relevance impacts the processing speed of emotional expressions.

Many further questions arise from these conclusions. For example, if expressions are decodable within 100 ms, how does presentation duration impact these pattern differences? Are expressions decodable outside awareness? How does behavioural relevance impact these temporal dynamics when expression itself is the object of behavioural goals, such as during an expression recognition task? These questions are addressed in the next chapter of this thesis (Chapter 3), which employs rapid presentation of emotional expressions and an expression discrimination task to interrogate the neural representations of faces in the presence of limited visual input.

Chapter 3

Configural representations support rapid face perception

3.1 Abstract

In the previous chapter, we focused on rapid implicit processing of emotional faces. Here, we turn to an expression recognition paradigm in order to explore the link between facial features, brain and behaviour. To investigate how this relationship changes in challenging viewing conditions and outside awareness, we varied the presentation duration of backward-masked facial expressions. The results indicated that face perception was supported by a two-stage process, with the ventral stream encoding facial features at an early stage and facial configuration at a later stage. Reducing presentation time modulated this process: early responses were transient, while featural and configural representations emerged later. These patterns overlapped with representations of behaviour in ventral stream areas, pointing to their importance in extracting task-relevant information. Although both face presence and expression were decodable from MEG data when stimuli were presented as briefly as 30 ms, only face presence could be decoded outside of subjective awareness. These results highlight the efficient feature extraction performed in the visual system in order to support rapid face categorization.

3.2 Introduction

Behavioural goals are thought to heavily influence how we process and perceive the world (Corbetta and Shulman, 2002; Gilbert and Sigman, 2007). Previous research has highlighted the role of task goals in shaping object and scene processing in the visual system (e.g. De Cesarei et al., 2018; Groen et al., 2018). Similarly, differences in how emotional faces are processed in passive viewing, as opposed to when they are the object of a task, have been frequently shown (Frühholz et al., 2011; Kliemann et al., 2016; Krolak-Salmon et al., 2001; Lange et al., 2003).

In this chapter, we move from the passive viewing paradigm discussed in Chapter 2 to a task involving explicit expression recognition. By manipulating the presentation duration of emotional face stimuli, we address three questions: (1) How are emotional faces processed in challenging viewing conditions? (2) Are emotional faces processed outside of subjective awareness? (3) How are behaviour and face features represented in MEG responses?

The rapid, bottom-up processing of emotional expressions is thought to extend to unconscious processing, although the extent and mechanisms of face perception outside of awareness are still not well understood. Using different methods of rendering faces "invisible", such as binocular suppression or backward masking, many experiments have shown some degree of unconscious face processing, demonstrated at the behavioural or neural levels (see Axelrod et al., 2015 for a review). However, electrophysiological investigations paint a complex picture of the underlying mechanisms: while many studies using binocular suppression have detected evoked responses to invisible faces (Jiang et al., 2009; Sterzer et al., 2009), other studies report no such effect, particularly when using backward masking (Fisch et al., 2010; Navajas et al., 2013; Reiss and Hoffman, 2007; Rodriguez et al., 2013), which is thought to disrupt re-entrant processing through conflicting input from feedforward connections (Lamme et al., 2002).

Facial expression has been shown to modulate the early stages of visual perception (Chapter 2) and to elicit non-conscious responses in numerous studies (Tamietto and De Gelder, 2010). Evidence of "blindsight" (non-conscious perception despite visual cortex lesions; e.g. Pegna et al., 2005) has led to considerable debate

about the automaticity of emotion perception and the role of a subcortical route in facilitating it (Pessoa, 2005; Pessoa et al., 2005b), with most of the evidence showing a processing advantage for invisible fearful faces (e.g. Bertini et al., 2017; Jiang and He, 2006; Williams et al., 2004). However, some studies show evidence against the non-conscious processing of expression (Hedger et al., 2016; Schlossmacher et al., 2017). Furthermore, it is unclear whether the advantage for fearful faces found in many experiments generalizes to threat-related expressions, or is linked to characteristic low-level properties (Hedger et al., 2015). This idea is reinforced by inconsistent effects found for angry faces: while some experiments show evidence of non-conscious perception of angry faces (e.g. Adams et al., 2010; Almeida et al., 2013), other studies show no effect or even a disadvantage in the competition for awareness (Gray et al., 2013; Hedger et al., 2015, 2016).

In addition, although evidence of rapid face processing points to highly efficient feature extraction, the mechanisms supporting this are still the subject of debate. It is widely believed that faces are perceived holistically, unlike other stimuli (Farah et al., 1998; Richler and Gauthier, 2014); however, some behavioural goals, such as identity recognition, are thought to rely on facial features and not on holistic perception (Visconti Di Oleggio Castello et al., 2017). Classic models support a configural model of face perception (Calder et al., 2000; Namdar et al., 2015), from the detection of a first-order configuration (face features) to the perception of a second-order configuration determined by relationships between features (Maurer et al., 2002; Piepers and Robbins, 2012). Although classic paradigms like face inversion or the composite face have shown how the highly specialized mechanisms for face perception can break down in the presence of configural disruption (Behrmann et al., 2014), the spatiotemporal dynamics of these processes remain less well understood.

In this chapter, we varied stimulus duration to interrogate the neural representations underpinning rapid face and expression perception, and we tracked how they change in the presence of limited visual input. We then used multivariate methods to assess the presence of neural responses to faces presented outside of subjective awareness. We reliably detected a neural response to subliminal faces, but no expression modulation outside of awareness, although expression

contributed to behavioural responses to invisible faces. Finally, we used representational similarity analysis (RSA) to tease apart the contributions of first-order and second-order face configuration and to explore the link between behaviour and ventral stream responses to faces. Together, these analyses highlight a face processing system highly adaptable to both behavioural goals and challenging viewing conditions.

3.3 Materials and Methods

3.3.1 Participants

The participants were 25 healthy volunteers (16 female, age range 19-42, mean age 25.6 ± 5.39). All volunteers gave written consent to participate in the study in accordance with The Code of Ethics of the World Medical Association (Declaration of Helsinki). All procedures were approved by the ethics committee of the School of Psychology, Cardiff University.

3.3.2 Stimuli

Stimuli were 20 faces with angry, neutral and happy expressions (10 female faces) from the NimStim database (Tottenham et al., 2009). The eyes were aligned across all faces using automated eye detection as implemented in the Matlab Computer Vision System toolbox. An oval mask was used to crop the faces to a size of 378×252 pixels subtending 3.9×2.6 degrees of visual angle. All images were converted to grayscale. Their spatial frequency was matched by specifying the rotational average of the Fourier amplitude spectra as implemented in the SHINE toolbox (Willenbockel et al., 2010), and Fourier amplitude spectra for all faces were set to the average across the face set.

Masks and control stimuli were created by scrambling the phase of all face images in the Fourier domain (Perry and Singh, 2014). To ensure matched low-level properties between face and control stimuli, pixel intensities were normalized between each image and its scrambled counterpart, using the minimum and maximum pixel intensity of the scrambled image.

3.3.3 Experimental design

At the start of each trial, a white fixation cross was centrally presented on an isoluminant gray background. Its duration was pseudorandomly chosen from a uniform distribution between 1.3 and 1.6 s. A face stimulus was then centrally presented with a duration of either 10 ms, 30 ms or 150 ms; the stimulus was followed by a phase-scrambled mask with a duration of 190 ms, 170 ms or 50 ms respectively (for a constant total stimulus duration of 200 ms). In each block, 10 trials contained no face; instead, a phase-scrambled control stimulus was flashed for 10 ms and followed by another mask.

After a 500 ms delay intended to dissociate face perception from response preparation, participants had to correctly select the expression they had perceived out of three alternatives presented on screen (Figure 3.3A). They had 1.5 seconds to make a button press; if they were sure that no face had been presented, they could refrain from responding. The mapping of the response buttons to emotional expressions changed halfway through the experiment so as to ensure that emotional expression processing would not be confounded by specific motor preparation effects.

Next, participants had to rate how clearly they had seen the face using a 3-point scale starting from 0. They were instructed to only select 0 if no face had been perceived, 1 if they had perceived a face but not clearly, and 2 if they had clearly perceived the face. They had 2 seconds to make this response.

In each of four blocks, each face was presented once with each of the three possible stimulus durations. We thus collected 80 trials per condition, except for the control condition (containing scrambled faces) which only had 40 trials.

3.3.4 Data acquisition

All participants with one exception acquired a whole-head structural MRI on a 3T General Electric or Siemens scanner using a 1 mm isotropic Fast Spoiled Gradient-Recalled-Echo pulse sequence.

Whole-head MEG recordings were made using a 275-channel CTF axial gradiometer system at a sampling rate of 1200 Hz. Four of the sensors were turned off due to excessive sensor noise. An additional 29 reference channels were recorded

for noise rejection purposes and the primary sensors were analyzed as synthetic third-order gradiometers (Vrba and Robinson, 2001).

Stimuli were presented using a ProPixx projector system (VPixx Technologies) with a refresh rate set to 100 Hz. Images were projected to a screen with a resolution of 1920 x 1080 pixels situated at a distance of 1.2 m from the participant. Recordings were made in four blocks of approximately 15 minutes each, separated by short breaks. The data were collected in 2.5 s epochs beginning 1 s prior to stimulus onset.

Participants performed the task while sitting upright. To continuously monitor head position relative to a fixed coordinate system on the dewar, electromagnetic coils were attached to the nasion and pre-auricular points on the participants' scalp. To help co-register the MEG data with the participants' structural MRI scans, the head shape of each subject was defined using an ANT Xensor digitizer (ANT Neuro). An Eyelink 1000 eye-tracker system (SR Research) with a sampling rate of 1000 Hz was used to track the subjects' right pupil and corneal reflex.

3.3.5 Behavioural analysis

The effect of stimulus duration and emotional expression on participants' expression discrimination accuracy (percentage correct responses) was analyzed after applying a rationalized arcsine transformation (Studebaker, 1985) using a 3x3 repeated-measures ANOVA with factors *Duration* (levels: 10 ms, 30 ms, and 150 ms) and *Expression* (levels: angry, happy, and neutral).

3.3.6 Event-related field analysis

We assessed the presence of differences between conditions in event-related fields (ERF). For the purposes of this analysis, MEG data were bandpass-filtered between 0.1 and 30 Hz and axial gradiometer event-related fields were averaged across subjects to calculate the global field power across all trials and conditions. This allowed us to determine three time windows of interest for evoked response component analysis: 63-137 ms (M100), 137-203 ms (M170), and 203 – 306 ms (M220).

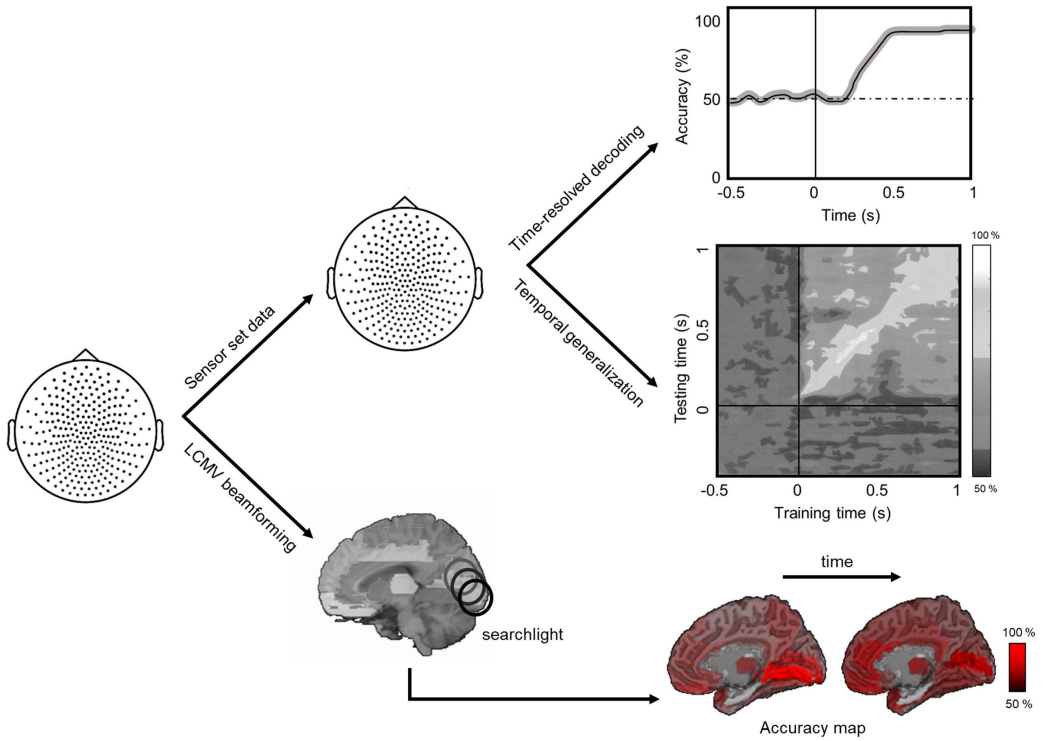

FIGURE 3.1: MPA analysis framework used in this chapter. Temporal dynamics were assessed using time-resolved and cross-time (temporal generalization) decoding at the sensor level, while spatial information was investigated using a searchlight approach in source space.

Next, we averaged evoked response fields for each condition and subject within the three time windows. We tested for differences between responses to faces and scrambled stimuli, and between responses to different emotional expressions, using paired t-tests and repeated-measures ANOVAs respectively at each sensor and time window. Significant sensors were determined using randomization testing (5000 iterations) and corrected for multiple comparisons using the maximal statistic distribution ($\alpha = 0.001$ to correct for multiple tests).

3.3.7 MEG multivariate pattern analysis (MVPA)

To test for differences between conditions present in multivariate patterns, we used a linear Support Vector Machine (SVM) classifier with L2 regularization and a box constraint $c = 1$. The classifier was implemented in Matlab using LibLinear (Fan et al., 2008) and the Statistics and Machine Learning Toolbox (Mathworks, Inc.).

We performed binary classification on (1) responses to neutral faces versus scrambled stimuli (face decoding); (2) all three pairs of emotional expressions (expression decoding).

For face decoding, time-resolved classification was performed separately for each stimulus duration (Figure 3.1). To assess the presence of subjectively nonconscious responses, the classification of faces presented for 10 ms was performed after excluding any trials reported as containing a face. To ensure that decoding results were not biased by stimulus repetitions or recognition of face identities across the training and test sets, cross-exemplar five-fold cross-validation was used to assess classification performance: the classifier was trained on 16 of the 20 face identities and 8 of the 10 scrambled images, and tested on the remaining 4 faces and 2 scrambled exemplars.

To assess similarities between responses across stimulus duration conditions, face cross-decoding was also performed, whereby a decoder was trained on 150 ms faces and tested on 30 ms faces and viceversa. The analysis was repeated for all pairs of conditions, using cross-exemplar cross-validation to ensure true generalization of responses; the resulting accuracies were averaged across the two training/testing directions, which led to similar results.

The temporal structure of face-related information was assessed through temporal generalization decoding (King and Dehaene, 2014). Classifier models were trained on each sampled time point between -0.1 and 0.7 s and tested on all time points in order to evaluate the generalizability of neural patterns over time at each stimulus duration. For this analysis, a cross-exemplar hold-out procedure was used to speed up computation (the training and test sets each consisted of 10 face identities/5 scrambled exemplars).

For expression decoding, classification was separately applied to all pairs of emotional expression conditions for each stimulus duration and perceptual awareness rating. As low trial numbers were a limitation of the study design, we increased the power of our analysis by also pooling together trials containing faces shown for 30 ms and 150 ms (which were shown to share representations in the cross-decoding analysis). Performance was evaluated using five-fold cross-exemplar cross-validation.

To achieve equal class sizes in face decoding, face trials were randomly subsampled (after cross-exemplar partitioning) to match the number of scrambled trials. For expression classification, trial numbers did not significantly differ between conditions after artefact rejection ($F(1.92, 46.18) = 0.15, P = 0.85, \eta^2 = 0.0062$).

3.3.8 MEG sensor-level analyses

MEG data were analyzed using Matlab and the Fieldtrip toolbox. Prior to analysis, trials containing excessive eye or muscle artefacts were excluded based on visual inspection, as were trials exceeding 5 mm in head motion (quantified as the displacement of any head coil between two sampled time points). Using eyetracker information, we also excluded trials containing saccades and fixations away from stimulus or blinks during stimulus presentation. A mean of 8.71% ±9.4% of trials were excluded based on this procedure.

For all analyses, MEG data were downsampled to 300 Hz and baseline corrected using the 500 ms before stimulus onset. A low-pass filter was applied at 100 Hz and a 50 Hz comb filter was used to remove the mains noise and its harmonics.

To improve SNR (Grootswagers et al., 2017), each dataset was divided into 20 equal partitions and pseudo-trials were created by averaging the trials in each partition. This procedure was repeated 10 times with random assignment of trials to pseudo-trials and was performed separately for the training and test sets.

To improve data quality, we performed multivariate noise normalization (MNN; Guggenmos et al., 2018). The time-resolved error covariance between sensors was calculated based on the covariance matrix (Σ) of the training set (X) and used to normalize both the training and test sets, in order to downweight MEG channels with higher noise levels (Equation 3.1).

$$X^* = \Sigma^{-\frac{1}{2}} X \tag{3.1}$$

In sensor-level MVPA analyses, all 271 MEG sensors were included as features and decoding was performed for each sampled time point between -0.1 and 0.7 s around stimulus onset.

3.3.9 MEG source-space analyses

For source analyses, participants' MRI was coregistered to the MEG data by marking the fiducial coil locations on the MRI and aligning the digitized head shape to the MRI with Fieldtrip. Note that the participant who had not acquired an MRI was excluded from source-space analyses. MEG data were projected into source space using a vectorial LCMV beamformer (Van Veen et al., 1997). To reconstruct activity at locations equivalent across participants, a template grid with a 10 mm isotropic resolution was defined using the MNI template brain and was warped to each participant's anatomical MRI. The covariance matrix was calculated based on the average of all trials across conditions bandpass-filtered between 0.1 and 100 Hz; this was then combined with a single-shell forward model to create an adaptive spatial filter, reconstructing each source as a weighted sum of all MEG sensor signals (Hillebrand et al., 2005). To alleviate the depth bias in MEG source reconstruction, beamformer weights were normalized by their vector norm (Hillebrand et al., 2012).

To improve data quality, MNN was included in the source localization procedure. As beamforming constructs a common filter based on pooled data (thus introducing no condition-related bias), the error covariance was in this case also calculated based on the pooled data. We then multiplied the normalized beamformer filters by the error covariance matrix, ensuring that the filters downweighted sensors with higher noise levels. The time-courses of virtual sensors were then reconstructed at all locations in the brain by multiplying the sensor-level data by the corresponding weighted filters. This resulted in three time-courses for each source, containing each of the three dipole orientations, which were concatenated for use in the MVPA analysis in order to maximize classification performance (Gohel et al., 2018). Preprocessing (baseline correction and downsampling) was performed as for sensor-level analyses.

A searchlight approach was used in source-space classification, whereby clusters with a 10 mm radius were entered separately into the decoding analysis. To

exclude sources outside the brain and in regions such as the cerebellum, we restricted our searchlight analysis to 1256 sources included in the 90-region Automated Anatomical Clustering (AAL) atlas (Tzourio-Mazoyer et al., 2002). Given the 10 mm resolution of our sourcemodel, this amounted to a maximum of 27 neighbouring sources being included as features (mean 26.9, median 27, SD 0.31). Decoding of subliminal faces vs. scrambled stimuli was performed on 30 ms time windows with a 3 ms overlap using the time windows identified in sensor-space decoding in order to reduce computational cost.

We also performed supraliminal face decoding (150 ms faces vs. scrambled stimuli) in order to identify a face-responsive ROI for use in the RSA analysis. This was accomplished by identifying searchlights achieving a cross-subject accuracy above the 99.5th percentile (P<0.005, 66 searchlights; Figure 3.2). To assess whether this area also encoded expression-related information, source-space decoding of expression was performed using a searchlight approach within this ROI.

3.3.10 Significance testing

We evaluated decoding performance using the averaged accuracy across subjects (proportion correctly classified trials) and assessed its significance through randomization testing (Jamalabadi et al., 2016; Nichols and Holmes, 2001; Noirhomme et al., 2014).

For sensor-level decoding, we repeated the cross-exemplar decoding procedure with 1,000 label shuffling iterations across the training and test sets. To speed up computation, the null distribution was estimated based on the time point achieving maximum overall accuracy in the MVPA analysis (Dima et al., 2018a). Observed time-resolved accuracies were then compared to the group maps to calculate P-values.

For whole-head sensor-space decoding, p-values were calculated using the maximal null distribution across tests (Nichols and Holmes, 2001; Singh et al., 2003) and corrected with a false discovery rate of 0.05, and a threshold of at least 5 consecutive significant time points was imposed. For temporal generalization decoding, the maximal distribution was created across tests and time points, and contiguous clusters of at least 5^2 time points were considered significant.

To detect above-chance decoding in source space, we performed 100 randomization iterations for each source cluster and subject in order to minimize computational cost. We then randomly combined the individual randomized accuracies into 10^3 whole-brain group maps (Stelzer et al., 2013). P-values were corrected across time using a FDR correction and a minimal extent of three consecutive time windows.

3.3.11 Representational Similarity Analysis (RSA)

Neural patterns and analysis framework

To interrogate the content of neural representations in space and time, we performed Representational Similarity Analysis (RSA). For this analysis, MEG data were source reconstructed as described above and trials were sorted according to expression and face identity. RSA was performed separately for each stimulus duration and only trials containing faces were included in the analysis.

To offset computational cost, a searchlight analysis was performed using occipitotemporal sources identified in face decoding, with a temporal resolution of 30 ms, as in the source-space decoding analysis. All three dipole orientations were concatenated for each source. The exclusion of responses to scrambled stimuli from the RSA ensured that feature selection was based on an orthogonal contrast (Figure 3.2).

To create MEG representational dissimilarity matrices (RDMs), we calculated the squared cross-validated Euclidean distance between all pairs of face stimuli (Guggenmos et al., 2018). Note that as the data were multivariately noise- normalized, this is equivalent to the squared cross-validated Mahalanobis distance (Walther et al., 2016). For each participant, the data were split into a training set (the first 2 sessions) and a test set (the last 2 sessions). The two stimulus repetitions contained in each set were averaged, and these were averaged across subjects to create training and test sets. To compute the cross-validated Euclidean distance between two stimulus patterns (X^*, Y^*), we calculated the dot products of pattern differences based on the training set and the test set (Equation 3.2). This procedure

has the advantage of increasing the reliability of distance estimates in the presence of noise.

$$d^2(X^*, Y^*) = \sum_{i=1}^{n} (X_i^* - Y_i^*)_{train} (X_i^* - Y_i^*)_{test} \tag{3.2}$$

The spatiotemporally resolved MEG RDMs were then correlated with several model RDMs to assess the contribution of different features to neural representations. In an initial analysis, we calculated Spearman's rank correlation coefficients between each model RDM and the MEG RDM (Nili et al., 2014). To further investigate the unique contribution of each model, we entered the significantly correlated models based on visual features of the images into a partial correlation analysis, where each model's correlation to the MEG data was recalculated after partialling out the contribution of the other models.

Note that a model based on behaviour, which was also represented in the MEG data for all stimulus duration conditions, was not included in the partial correlation analysis; the rationale is that we were interested in the contribution of each visual property independently of the others, but we did not expect a unique contribution of behaviour in the absence of expression-related visual properties, and partialling out the behavioural model from the visual models would not be easily interpretable. Instead, we preferred to independently describe the correlations between behaviour and visual models, brain and behaviour, and brain and visual models, as the three main factors of interest in our analysis.

Model RDMs

We investigated the temporal dynamics of face perception by assessing the similarity between MEG patterns and 9 models quantifying behaviour and facial/visual properties (Figure 3.2).

To create behavioural model RDMs, we calculated the number of error responses made by each participant to each stimulus and summed these up to create a cross-subject behavioural RDM. For each stimulus duration, we created separate behavioural RDMs by calculating pairwise cross-validated Euclidean distances between error response patterns, using a cross-session training/test split as described

TABLE 3.1: Action Units used to create a model RDM

AU Code	Facial Action Coding System Name
AU01	Inner brow raiser
AU02	Outer brow raiser
AU04	Brow lowerer
AU06	Cheek raiser
AU09	Nose wrinkler
AU10	Upper lip raiser
AU12	Lip corner puller
AU14	Dimpler
AU15	Lip corner depressor
AU17	Chin raiser
AU20	Lip stretcher
AU25	Lips part

above.

To create a high-level identity model, we assigned distances of 0 to pairs of face identities repeated across emotional expression conditions, and distances of 1 to pairs of different face identities. We used a similar strategy to create high-level emotional expression models. An all-versus-all model was created by assigning distances of 0 to all faces belonging to the same emotional expression condition, and distances of 1 to pairs of faces differing in emotion. We also tested a neutral-versus-others model by assigning distances of 0 to all emotional faces (happy + angry), and an angry-versus-others model by assigning distances of 0 to all benign faces (happy + neutral).

To account for variability in expression that is not captured by such high-level binary representations, we also tested a model based on Action Units. Action Units quantify changes in expression by categorizing facial movements (Ekman and Friesen, 1977). We used OpenFace (Baltrusaitis et al., 2016) to automatically extract the intensity of 12 Action Units in our image set (Table 3.1), and we calculated pairwise Euclidean distances between these intensities for all pairs of faces in our stimulus set to obtain an Action Unit RDM.

To create face configuration RDMs, we also used OpenFace (Baltrusaitis et al., 2016) to automatically detect and label face landmarks. The software created 68 2D landmarks for each face. We removed landmarks corresponding to the face outline and the 2 outermost eyebrow landmarks, to account for cases in which these

landmarks were cropped out by the oval mask used in the MEG stimulus set. The final landmark set consisted of 47 coordinates for 6 facial features (eyes, eyebrows, nose, and mouth), which were visually inspected to ensure that they were correctly marked. To capture feature-based (local) facial configuration, we calculated within-feature pairwise Euclidean distances between landmarks (Figure 3.2C). To quantify global face configuration, we calculated between-feature Euclidean distances (the distances between each landmark and all landmarks belonging to different facial features). Distances were then concatenated to create feature vectors describing each face in terms of its local/global configuration, and Euclidean distances between them gave the final configural model RDMs. The local/global configurations correspond to the first-order (isolated) and second-order (relational) features in classic configural models of face perception (Diamond and Carey, 1986; Piepers and Robbins, 2012).

Finally, a spatial envelope model was created in order to capture image characteristics using the GIST descriptor (Oliva and Torralba, 2001). This procedure extracted 512 values per image by applying a series of Gabor filters at different orientations and positions, and thus quantified the average orientation energy at each spatial frequency. To obtain the spatial envelope RDM, we calculated pairwise Euclidean distances between all images using the GIST values.

Significance testing

To assess the significance of spatiotemporally resolved correlation maps, we used a randomization approach (3.3.10). Model RDMs were shuffled 1,000 times and correlations were recomputed for each of the 66 searchlights using the time window achieving the maximal correlation coefficient across models for each of the stimulus duration conditions. Since negative correlations were not expected and would not be easily interpretable, P-values were calculated using a one-sided test (Furl et al., 2017). To correct for multiple comparisons, P-values were omnibus-corrected by creating a maximal distribution of randomized correlation coefficients across searchlights, models and conditions, and FDR and cluster-corrected across timepoints ($\alpha = 0.05$, thresholded at 3 consecutive time windows).

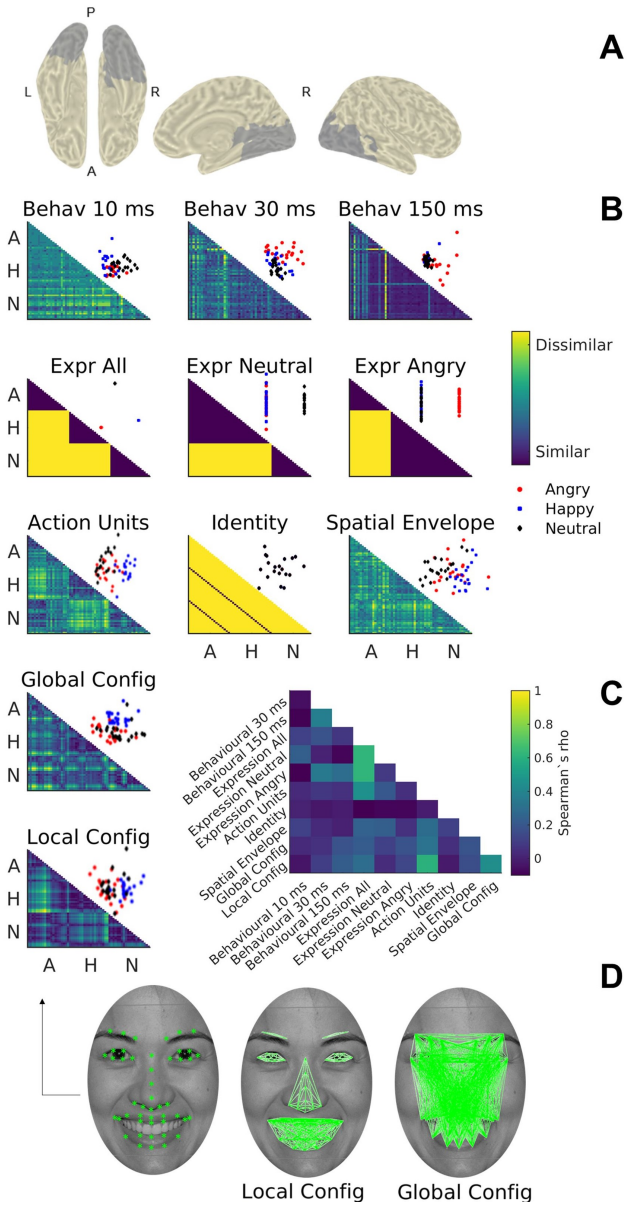

FIGURE 3.2: RSA models. Models used in RSA analysis. **A**. Sources included in the representational similarity analysis based on face vs. scrambled classification results. P: posterior; A: anterior; L: left; R: right. **B**. Model RDMs showing predicted distances between all pairs of stimuli (lower triangles). A:Angry; H: Happy; N: neutral. Stimuli are sorted according to face identity. Upper triangles show 2D multidimensional scaling (MDS) plots for each model, which help visualize the distances between stimuli according to each model. **C**. Model inter-correlations (Spearman's ρ). **D**. Metrics used to derive the local and global face configuration models. The left-hand panel shows automatically detected facial landmarks for an example stimulus, while the other two panels depict the pairwise Euclidean distances used to calculate the two model RDMs. Behav: behavioural models; Expr: high-level expression models (all-vs-all, neutral-vs-others, and angry-vs-others); Config: face configuration models.

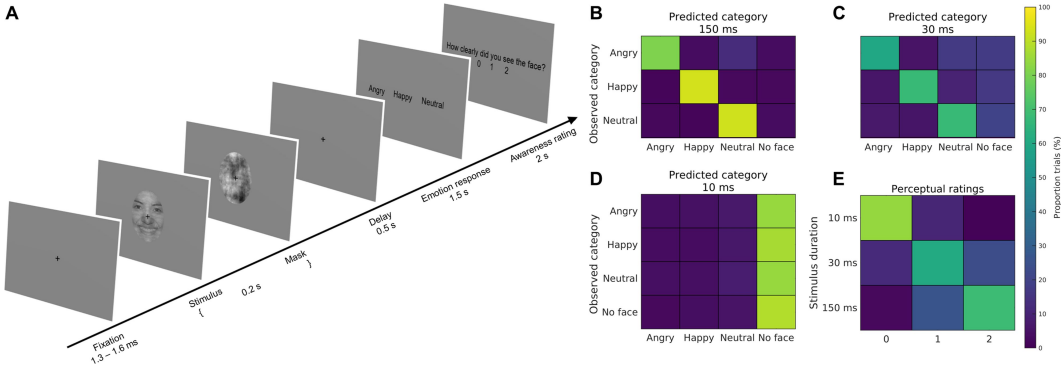

FIGURE 3.3: Overview of the experimental paradigm and behavioural results. **A**. Stimuli were presented on screen for 150 ms, 30 ms, or 10 ms, and were followed by a 50 ms, 170 ms, or 190 ms scrambled mask. **B-D**. Confusion matrices mapping the average proportion of trials receiving each of the possible responses (X-axis) out of the trials belonging to each category (Y-axis). "No response" trials were excluded for statistical analysis, but are shown here as representing a "no face" (i.e. scrambled face) response. Note that scrambled faces were only presented in the 10 ms condition. **E**. Perceptual ratings for each stimulus duration summarized as average proportion of trials.

Variance partitioning

To gain more insight into the relationship between behavioural responses, expression categories and face configuration models, we used a variance partitioning approach (Greene et al., 2016; Groen et al., 2018). For each stimulus duration condition, the corresponding behavioural RDM was entered into a hierarchical multiple linear regression analysis, with three model RDMs as predictors: the two facial configuration models and the most correlated high-level expression model (10 ms: neutral-vs-others; 30 and 150 ms: angry-vs-others). These models were selected to reduce the predictor space before performing variance partitioning. To quantify the unique and shared variance contributed by each model, we calculated the R^2 value for every combination of predictors (i.e. all three models together, each pair of models separately, and each model separately). The EulerAPE software was used for visualization (Micallef and Rodgers, 2014; Figure 3.2).

3.4 Results

3.4.1 Perception and behaviour

In order to assess the effects of stimulus duration and face expression on behaviour, we calculated confusion matrices mapping the expression discrimination responses to each stimulus category (Figure 3.3). We then performed a 3×3 repeated-measures ANOVA with factors *Duration* (levels: 10 ms, 30 ms, 150 ms) and *Expression* (levels: angry, happy, neutral). As expected, stimulus duration had a strong effect on expression discrimination performance, with average performance not exceeding chance level at 10 ms ($33.45\% \pm 2.99$) and rising well above chance at 30 and 150 ms ($78.62\% \pm 2.11$ and $91.83\% \pm 1$ respectively). This was reflected in a significant main effect of duration in the ANOVA ($P < 0.0001$, $F(1.21, 29.06) = 221.05$, $\eta^2 = 0.9$). Face expression had a weak effect, with angry faces categorized less accurately than both happy and neutral faces ($P = 0.046, F(1.95, 46.71) = 3.33, \eta^2 = 0.12$), with no significant interaction effect ($P = 0.23$, $F(1.74, 41.83) = 1.53$, $\eta^2 = 0.06$).

Participants found the task challenging, as reflected in the perceptual awareness ratings: 84.5% of the 10 ms trials were rated as not containing a face (Figure 3.3E). This suggests that participants were complying with the task with respect to both expression discrimination and perceptual rating. Importantly, for faces presented for 10 ms, there was no difference in accuracy between expressions ($P = 0.43, F(1.65, 39.5) = 0.8$) or between any pair of cells in the confusion matrix ($P = 0.6, F(3.42, 82.07) = 0.64$), suggesting that faces presented at this duration were equally likely to be categorized as any expression. Note that the expression discrimination task here was not a forced-choice task (participants could refrain from responding) and these tests were performed on the small subset of 10 ms trials that received a response; references to awareness in this chapter thus refer exclusively to subjective awareness, as indicated by perceptual ratings.

3.4.2 Evoked responses to faces

We assessed the presence of a response to faces by contrasting neutral faces with scrambled stimuli at each stimulus duration (Figure 3.4). For 150 ms faces, we found significant differences at M170 latencies and M220 latencies ($P < 0.0007, t(24) >$

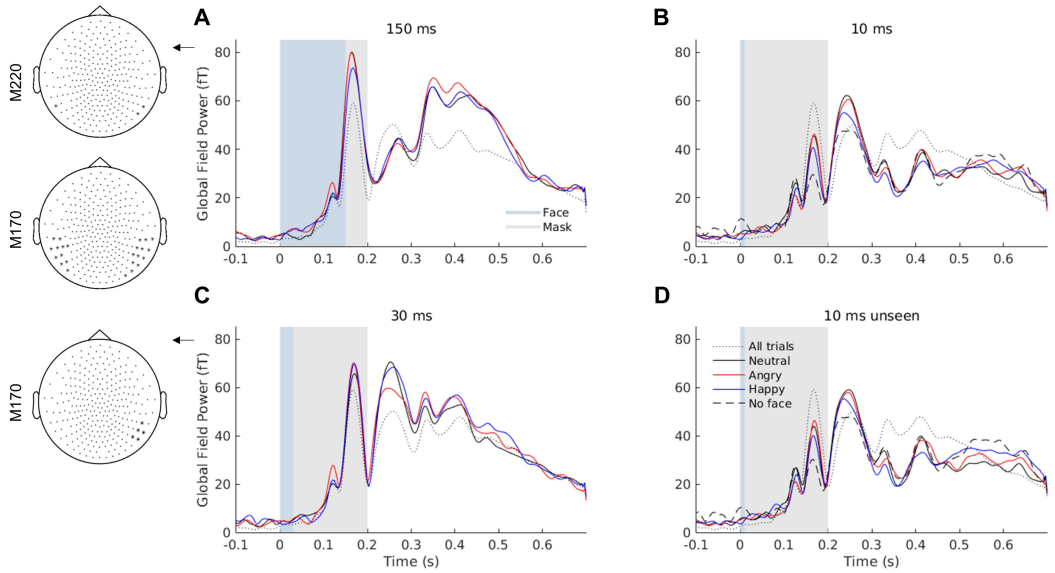

FIGURE 3.4: ERF analysis results. **A-D**. Global field power averaged across participants and trials for each stimulus duration condition. Note decreasing M170 amplitudes with stimulus duration. **Left**. Significant sensors in the face vs scrambled (no face) contrast at M170 (137-203 ms) and M220 (203-306 ms) latencies (P<0.001 corrected).

6.07), but no significant effects at M100 latencies surviving correction for multiple comparisons. A significant, but smaller, cluster of right temporal sensors was also found for 30 ms faces at M170 latencies ($P < 0.0004, t(24) > 5.99$). No conclusive effects were found when contrasting faces presented for 10 ms with their scrambled counterparts, regardless of whether trials where a face was perceived were excluded or not ($P > 0.015, t(24) < 4.66$ across comparisons), and no effect of emotional expression was found at any of the stimulus durations ($P > 0.06, F(2, 48) < 8.59$). Several factors could explain the absence of emotional expression effects in our ERF data: (1) stimuli were highly controlled for low-level properties, minimizing visually-driven differences in early time windows; (2) our time windows of interest did not include late stages dominated by task-related processing of expression; (3) we performed a whole-brain analysis with a conservative correction for multiple comparisons.

3.4.3 Spatiotemporal dynamics of face perception

To investigate face processing as a function of stimulus duration, we performed within-subject decoding of responses to faces vs. scrambled stimuli. The analysis

TABLE 3.2: Face decoding results

	Sensor-space			Source-space
	150 ms	**30 ms**	**10 ms**	**10 ms**
Max % accuracy	82.3	76.8	56.8	59.62
SD (%)	13.6	14.18	9.3	8.35
Decoding onset (ms)	100	100	147	120-150

included three components: sensor-level time-resolved classification to evaluate the progression of condition-related information; sensor-level temporal generalization to assess the temporal structure of this information; and source-space decoding to obtain spatial information about subliminal responses to faces (Figure 3.1).

Scrambled stimuli could be discriminated from faces presented for 150 and 30 ms as early as 100 ms, as reflected by above-chance decoding performance on the MEG sensor set (Figure 3.5A). After the initial peak in performance, decoding accuracy decreased, but remained well above chance for the remainder of the decoding time window. For faces presented for 10 ms and reported as not perceived, there was only a weak increase in decoding performance, which reached significance at 147 ms and dropped back to chance level after ~350 ms (Table 3.2).

To assess how well face representations generalized across stimulus durations, we repeated this analysis by training and testing on stimulus exemplars presented for different amounts of time (Figure 3.5B). Decoding accuracy was high when cross-decoding between 30 ms and 150 ms faces; interestingly, after an initial peak (100-200 ms), performance decreased, and started increasing again after 300 ms, suggesting that representations become more similar over time. On the other hand, representations only generalized to 10 ms faces for a limited time window, with a peak at M170 latencies.

Using temporal generalization decoding (King and Dehaene, 2014), we investigated the temporal structure underpinning face decoding, and we found that this changed with stimulus duration. For faces presented for 150 ms, successful temporal generalization started at ~93 ms in a diagonal pattern suggestive of transient representations, with more sustained representations (square patterns) arising at M170 latencies and after 300 ms (Figure 3.5D-E). For 30 ms stimuli, a diagonal

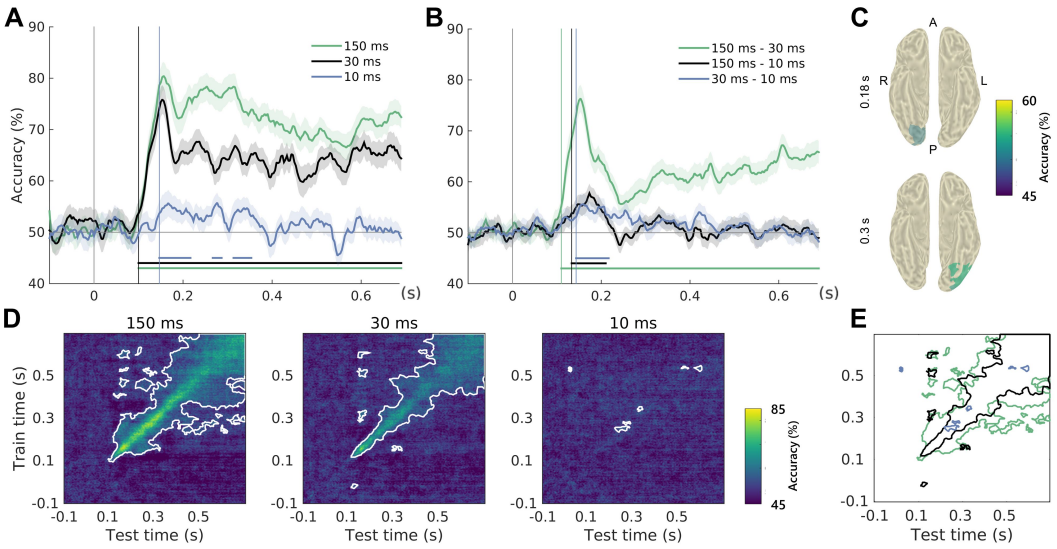

FIGURE 3.5: Face vs. scrambled decoding results. **A**. Sensor-space time-resolved decoding accuracy for all stimulus durations. Vertical bars mark above-chance decoding onset and horizontal lines show significant time windows (P<0.05, corrected). **B**. Sensor-space time-resolved cross-decoding for all pairs of stimulus durations. **C**. Sources achieving above-chance decoding of 10 ms faces outside awareness at M170 latencies (P<0.005, corrected). **D**. Sensor-space temporal generalization accuracy and significant clusters (white contours; P<0.05, corrected) for all stimulus durations. **E**. Significant temporal generalization clusters for all three stimulus durations, showing more sustained representations of faces presented for 150 ms (legend as in A).

generalization pattern started at ~110 ms after stimulus onset and sustained representations only arose later (~400 ms). Face processing thus appears to be heavily biased by stimulus presentation duration, with 30 ms faces failing to elicit a stable representation at M170 latencies. For faces presented for 10 ms, only few transient clusters survived correction for multiple comparisons, with the largest one occurring after 200 ms.

Finally, we spatially localized the subliminal response to faces in source space by performing whole-brain searchlight classification of 10 ms faces vs. scrambled stimuli (N=24). Faces were successfully decoded in a right occipital area at M170 latencies (Figure 3.5C), with a later stage associated with ventral patterns. Given the disruption of recurrent processing through backward masking in this paradigm, the occipital sources likely reflect the feedforward nature of this response.

TABLE 3.3: Sensor-space expression decoding results

	Stimulus duration											
	150 ms			30 ms			10 ms			30 + 150 ms		
	A-N	H-N	A-H	A-N	H-N	A-H	A-N	H-N	A-H	A-N	H-N	A-H
Max % accuracy	61.9	63.1	60.76	57.79	58.49	58.12	56.62	55.86	55.87	60.48	60.21	59.74
SD (%)	8.57	6.78	9.34	10.91	9.92	10.38	10.88	9.11	13.66	9.04	10.52	13.41
Decoding onset (ms)	180	113	220	437	120	633	N/A	N/A	N/A	107	113	117
	Perceptual rating											
	2			1			0			2 + 1		
	A-N	H-N	A-H	A-N	H-N	A-H	A-N	H-N	A-H	A-N	H-N	A-H
Max % accuracy	59.55	62.54	64.03	56.56	56.88	56.63	57.64	55.32	56.01	60.43	62.25	60.24
SD (%)	12.24	11.6	10.82	12.1	13.63	13.21	14.46	10.24	12.47	11.95	12.07	12.25
Decoding onset (ms)	230	113	523	307	120	130	N/A	N/A	N/A	220	113	127

3.4.4 Temporal dynamics of expression perception

We performed sensor-level time-resolved decoding of all pairs of emotional expressions separately for each stimulus duration. The highest decoding performance was achieved on late responses to expressions presented for 150 ms (Figure 3.6A). Expressions presented for 30 ms also achieved above-chance decoding, although these effects were more transient. We also performed this analysis on pooled datasets (faces presented for 30 and 150 ms), as the face cross-decoding analysis showed that responses generalized between these two categories (Figure 3.5B). Complementary results were obtained using the pooled datasets (faces presented for 30 and 150 ms), which revealed a multi-stage progression for all expressions, with transient early decoding at M100 latencies and an increasing accuracy in late time windows (Figure 3.6B). A source-space analysis revealed successful decoding of all three pairs of expressions in occipitotemporal cortex, although angry faces were associated with more sustained patterns in this ROI (Figure 3.7).

However, we found no above-chance performance when decoding 10 ms expressions. This finding is in line with other studies finding no evidence of expression processing outside awareness (Hedger et al., 2016; Koster et al., 2007; Pessoa et al., 2006), and we explore potential reasons for this result in the discussion.

The temporal generalization analysis supported these findings, showing that different stages entail different temporal dynamics: while early decoding was supported by limited diagonal clusters (suggestive of transient representations), relatively more sustained responses emerged in later time windows (300-500 ms). Stable representations emerged earlier when decoding angry vs neutral faces, as

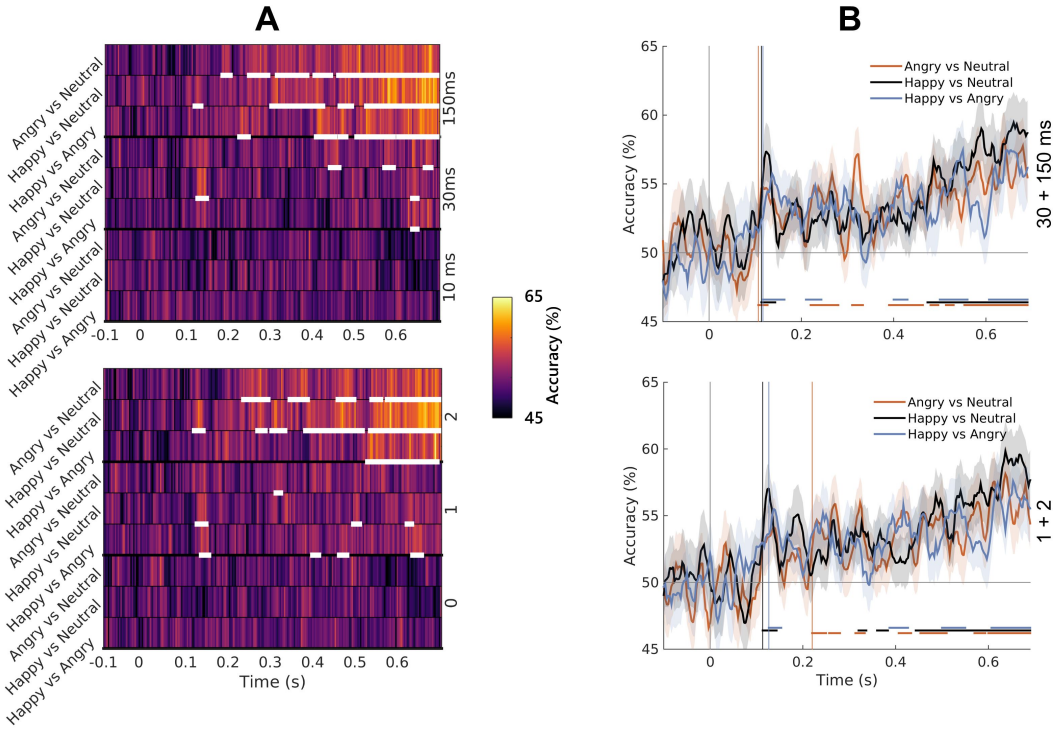

FIGURE 3.6: Expression decoding results. **A**. Time-resolved decoding accuracy for the three expression decoding problems and the three stimulus durations (above) / perceptual awareness ratings (below). White horizontal lines show significant time windows (P<0.05, corrected). **B**. Time-resolved accuracy for the three expression decoding problems using the pooled datasets (above: durations of 30 + 150 ms; below: perceptual ratings of 1 and 2).

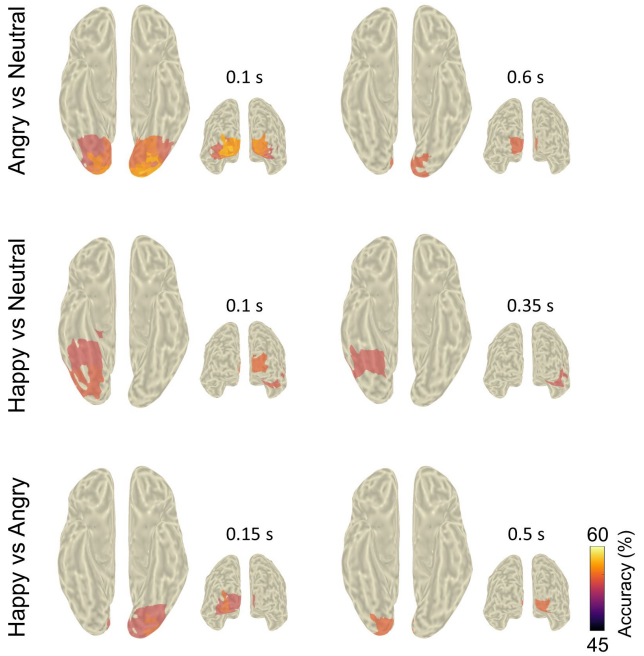

FIGURE 3.7: Source-space decoding of expression (pooled datasets) from searchlights in occipitotemporal cortex. Significant searchlights are plotted (P<0.05, corrected) at approximate onset and offset times.

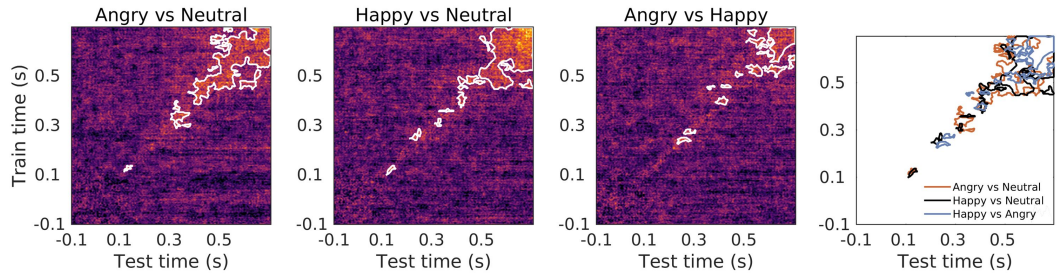

FIGURE 3.8: Temporal generalization patterns obtained using the pooled datasets (30 + 150 ms).

suggested by the earlier emergence of contiguous clusters (Figure 3.8).

3.4.5 Face representations in occipitotemporal cortex

To interrogate the content of neural representations in space and time, we performed representational similarity analysis (RSA) using a searchlight approach in face-responsive cortex at the source level (Su et al., 2012). We investigated the temporal dynamics of face perception by assessing the similarity between MEG patterns and models quantifying behaviour, expression, identity and visual properties.

Occipitotemporal cortex encodes behavioural responses

To assess the link between behaviour and neural patterns, we calculated model RDMs based on expression discrimination patterns across participants. Among the other model RDMs tested, behavioural RDMs correlated most with the high-level expression models (particularly the angry-vs-others model at 30 ms and 150 ms, Spearman's $\rho = 0.29$ and $\rho = 0.34$). At 150 ms, the behavioural RDM also correlated with the configural face models ($\rho = 0.22$ and $\rho = 0.18$). As expected based on performance, behavioural RDMs at 10 ms did not correlate with the other two ($\rho = -0.05$ and $\rho = -0.09$ respectively), while behavioural RDMs at 30 and 150 ms were positively correlated ($\rho = 0.38$; Figure 3.2B).

Based on these links, face configuration, together with facial expression, appears to partially explain behavioural responses. To more directly test this, we performed a variance partitioning analysis, using hierarchical multiple regression to quantify the unique and shared variance explained by facial configuration and

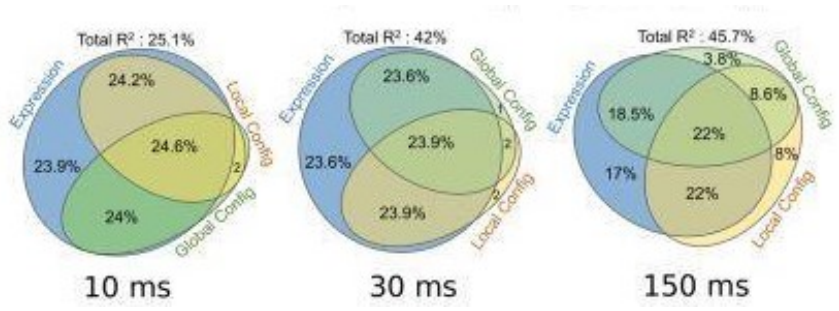

FIGURE 3.9: Variance partitioning results, showing the contributions of expression and face configuration models to behavioural responses at each stimulus duration. Values represent % of the total R^2.

high-level expression models in behavioural responses (3.3.11). In the 10 ms condition, the neutral-vs-others model and the two configural models explained 25.1% of the variance; in the 30 ms and 150 ms conditions, the angry-vs-others model and the configural models explained up to 45.7% of the variance in behaviour. Furthermore, while the expression model contributed most of the variance, over 75% of this variance was shared with the configural models. The unique contribution of configural models increased with stimulus duration (from ~2% at 10 ms, to ~20% at 150 ms). Together, these results point to the role of face configuration in driving high-level representations and behaviour. Note that for the 10 ms condition, we were unable to decode expression from the MEG data; however, expression and configuration explained a portion of the variance in behaviour, suggesting that they may contribute to the subliminal response to faces.

Behavioural RDMs showed the strongest and most sustained correlations with MEG patterns in ventral stream areas, including sources corresponding to the location of the fusiform face area (FFA) and OFA (Figure 3.10). Behavioural representations evolved differently in time for the three stimulus durations. For 10 ms faces, behaviour explained the data starting at 120 ms until the end of the analysis time window. Representations emerged at similar latencies for 150 ms faces and reached the noise ceiling before falling back to low ρ values at 400 ms. For 30 ms faces, correlations were significant starting at 210 ms in a relatively focal right temporal area. Patterns were more posterior for 10 ms faces and more extensive, including sources corresponding to the OFA and FFA, for 150 ms faces.

The correlation time-courses suggest interesting differences in processing as

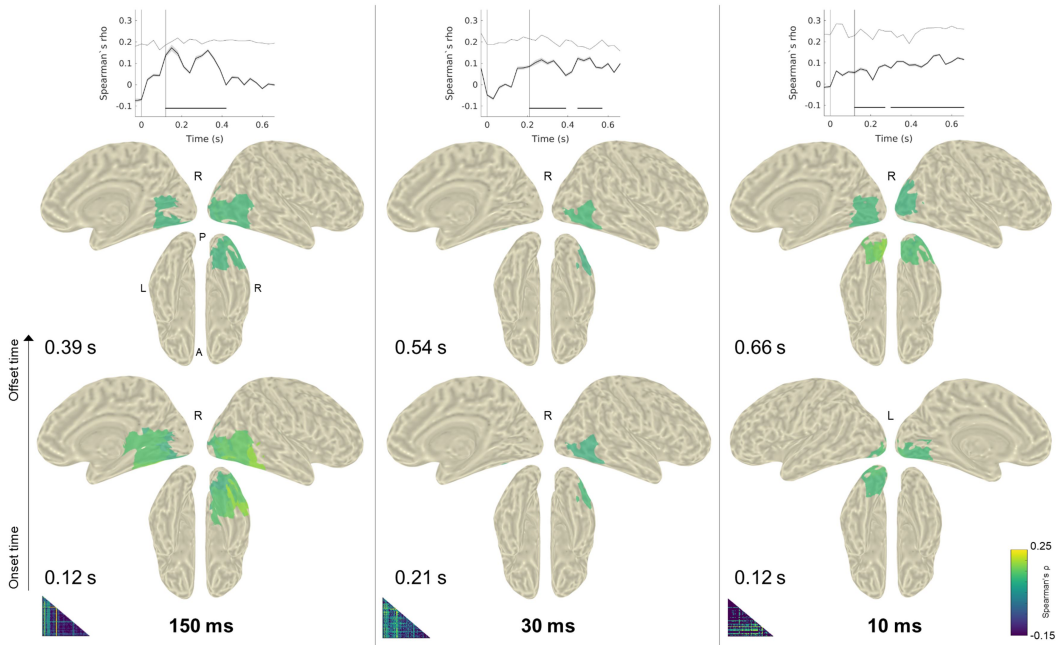

FIGURE 3.10: Correlations between MEG patterns and behavioural model RDMs for each stimulus condition duration (vertical columns). The top panels show correlation time-courses averaged across all significant searchlights; the noise ceiling is shown as a dotted horizontal line and is only approached in the 150 ms condition. The cortical maps show significant correlation coefficients for the first and last significant time windows (onset and offset times) on the inflated template MNI brain. The hemisphere shown is indicated with the letter R/L. Model RDMs are shown in the lower left corner of each column.

a function of the information available: for clearly perceived faces, features relevant in behaviour are extracted between 120-400 ms, while behavioural responses for briefly presented faces appear to require sustained processing, as reflected by behaviour-related correlations not dropping back to zero. These results are in line with previous evidence of behavioural representations in ventral stream areas in scene and object perception (e.g. Walther et al., 2009), and suggest that visual feature processing, even at relatively early stages, is closely linked to behavioural goals.

Configural face processing from featural to relational

The two face configuration models were also represented in the MEG patterns. In the correlation analysis, the local and global configuration models explained representations in partially overlapping areas of the ventral stream (corresponding to the right FFA location), with local configuration representations arising earlier (at

120 ms for 150 ms faces, and 300 ms for 30 ms faces). Our RSA method (3.3.11) favoured sustained correlations over transient peaks; note that the global configuration model correlation approached the noise ceiling during a transient time window at M170 latencies for both 150 ms and 30 ms faces, suggesting a contribution of second-order characteristics, although this occurred later than first-order feature representations (Figure 3.13). The partial correlation analysis revealed further differences between conditions: for 150 ms faces, the local and global models made unique contributions in explaining the data; conversely, for 30 ms faces we detected no unique contributions, suggesting that the extraction of configural information from faces occurs differently in the absence of sufficient information. None of the models significantly correlated with MEG patterns elicited by 10 ms faces.

Note that although both internal (eyes, nose, mouth) and external (face shape, hair) face features have been shown to contribute to neural responses to faces (Axelrod, 2010), we focus here on internal features; for the purposes of this paper, external features were excluded from the stimuli and we refer to the second-order configuration of distances between internal features as "global configuration". Internal features are relevant to the context of expression discrimination and have been shown to be more reliable even in facial recognition contexts (e.g. Kemp et al., 2016; Longmore et al., 2015).

Transient representations of visual and high-level models

Two other models elicited brief representations in the MEG data. For 150 ms faces, the spatial envelope model explained left hemisphere occipital representations starting at ~400 ms, suggesting sustained processing of visual features, potentially based on feedback mechanisms.

For 30 ms faces, a high-level expression model (neutral-vs.-others) was represented in the MEG data starting at 300 ms (Figure 3.12). This can be speculatively explained by the formation of task-related representations in the absence of sufficient information. Note that when faces are clearly presented, only specific facial feature models are represented, while categorical models show no contribution to occipitotemporal representations. On the contrary, when faces are briefly presented, the configural models do not contribute unique information, and only the

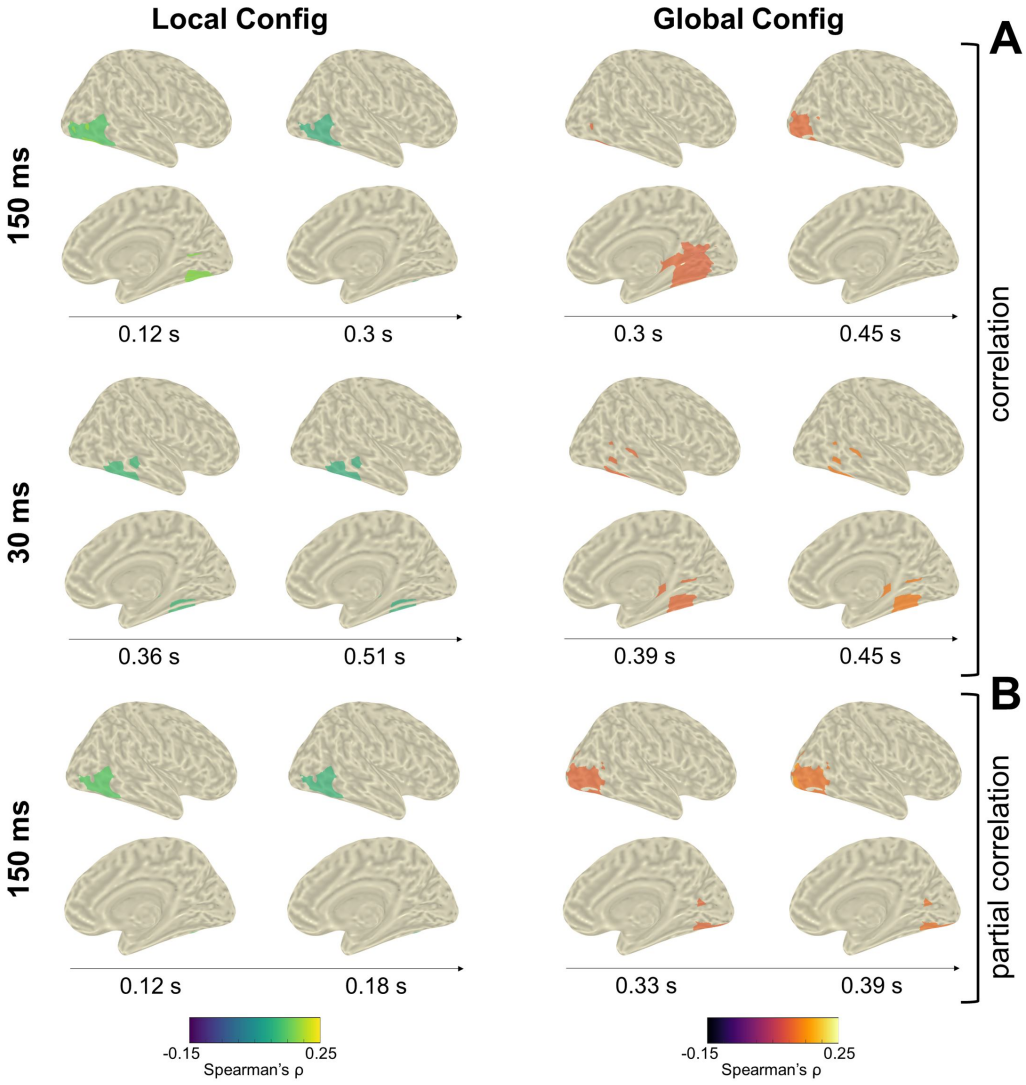

FIGURE 3.11: Significant correlations between MEG patterns and configural model RDMs. **A**: Correlation analysis results are significant for the 150 ms and 30 ms conditions. **B**: Partial correlation results are significant for the 150 ms condition. Only right hemisphere searchlights correlate with the configural models. Maps are shown for the onset and offset times of significant correlation.

high-level expression model is significant in the partial correlation analysis.

Although correlation coefficients between the models and neural data are generally low (maximum mean $\rho = 0.23$; Figure 3.13), the noise ceiling shows that the maximal correlation possible with our data is also low (mean $\rho = 0.21$); this is not surprising, considering the low ρ-values usually found in MEG RSA studies, and the fact that our paradigm involved complex, high-level visual stimuli and a demanding task. In this case, the noise ceiling serves as a useful benchmark for the explanatory power of our models. For example, the behavioural RDM reaches the noise ceiling in the 150 ms condition, but not for briefer stimuli, suggesting that behavioural representations fully explain the data when stimuli are clearly perceived. The local configuration model also shows good explanatory power at its earliest stage, and the same is true for the global model for a brief time window. With time, both models fall away from the noise ceiling, while other significant models also fail to fully explain the data (Figure 3.13).

Given the complex face processing and task-related activity reflected by the MEG patterns, it is not surprising that most models do not approach the noise ceiling. In fact, the explanatory power of the configural models at early stages (100-200 ms) is striking, as is the strength of behavioural representations in ventral stream within 400 ms. Furthermore, the initial peak in performance of the behavioural model overlaps with the peak of the local configuration model. Together with the shared variance between configuration, expression and behaviour shown in the variance partitioning analysis (Figure 3.9D), this points to the role played by facial configuration in the extraction of emotional cues essential in the expression discrimination task.

3.5 Discussion

In this chapter, we investigated how face representations in MEG sensor-level and source-space patterns vary with expression and with stimulus presentation duration. Using MVPA, we found a response to faces presented for 10 ms occurring at M170 latencies outside of subjective awareness, but no such response to expression. Furthermore, neural responses became more transient when presentation time was

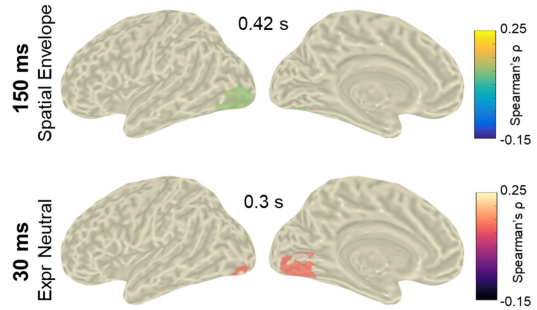

FIGURE 3.12: Significant correlations between: (1) MEG patterns for the 150 ms condition and the spatial envelope model RDM (**top**); (2) MEG patterns for the 30 ms condition and the high-level neutral-vs-others model (**bottom**). Only left hemisphere searchlights correlate with the two models. Maps are shown for the onset time of significant correlation, as clusters are sustained until offset (top: 0.54 s, bottom: 0.36 s).

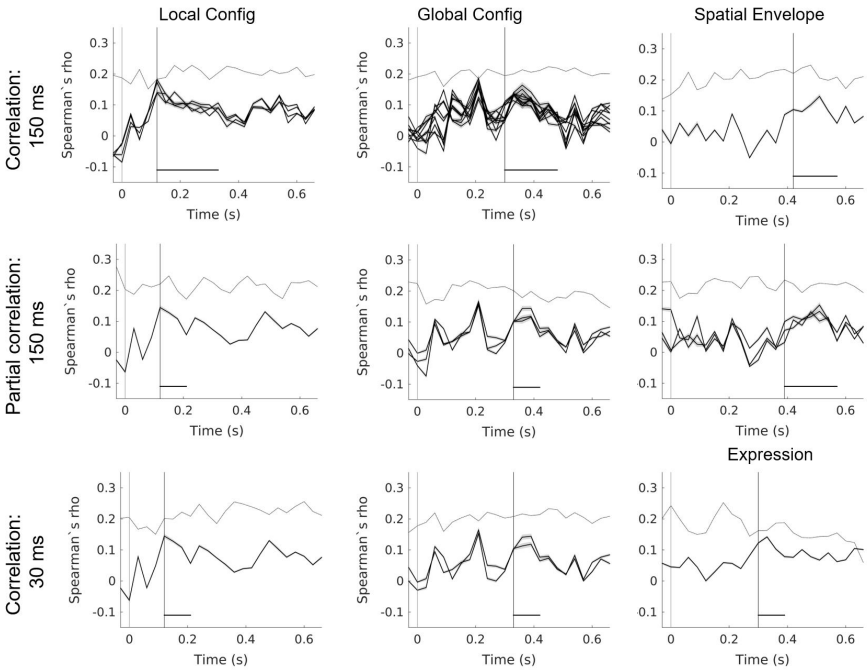

FIGURE 3.13: Correlation time-courses obtained in the RSA analysis. All significant searchlights are plotted separately against a noise ceiling averaged across significant searchlights.

reduced. Finally, we showed that behaviour and face configuration drive representations in face-responsive occipitotemporal cortex, with temporal dynamics varying as a function of stimulus duration.

3.5.1 Face and expression processing with limited visual input

When decoding faces and scrambled stimuli, we found early effects for 150 ms and 30 ms faces (~100 ms), as well as above-chance decoding of 10 ms faces shown outside of subjective awareness (140 - 350 ms), in line with previous studies showing evidence of face perception outside of awareness (Axelrod et al., 2015). Furthermore, temporal representations underpinning classification performance varied with stimulus duration: for 150 ms faces, a sustained representation emerged at M170 latencies which was absent for 30 ms faces. This suggests that clearly presented faces are perceived through a multi-stage process, while disrupted recurrent processing leads to delayed stable representations.

Conscious perception may be supported by temporally stable representations, while processing of stimuli outside subjective awareness may require a sequence of transient stages (Dehaene, 2016). Since above-chance decoding of 10 ms faces is transient in the current study, temporal generalization reveals only few transient clusters along the diagonal. On the other hand, the patterns differentiating 30 ms and 150 ms faces suggest that longer stimulus durations elicit an earlier stable representation, reflective of conscious perception and likely to be supported by recurrent processes. It has previously been suggested that faster stimulus presentation leads to more transient representations (Mohsenzadeh et al., 2018b); however, since the backward masking procedure used here disrupts the formation of a stable representation by entering the visual stream, it is unclear whether different methods of preventing awareness would lead to the same results.

Alternative explanations are possible when interpreting temporal generalization patterns. First, SNR decreases as a function of stimulus duration, and this could lead to lower accuracies and less sustained representations. However, we find that the most striking difference in temporal generalization patterns occurs at M170 latencies, which is the time window exhibiting comparable decoding accuracies between 150 ms and 30 ms faces. Thus, the transient patterns characterizing

M170 responses for rapidly presented faces are more likely to reflect a change in temporal dynamics. Second, it has been suggested that the transience of neural states can be overestimated in temporal generalization decoding due to trial-to-trial variability in effect onsets (Vidaurre et al., 2018); but since the conditions we are comparing differ only in stimulus duration, the progression from sustained to transient observed here is unlikely to be explained by differences in onset variability.

Information supporting face decoding outside of subjective awareness was localized mainly to occipital cortex in our searchlight source-space decoding analysis (Figure 3.5C). Given the suppression of sustained neural activity in backward masking, the early stages of this response can be attributed to either purely feedforward activity, or to feedback connections, which have been shown to target V1 at early stages of recurrent processing (Mohsenzadeh et al., 2018b; Wyatte et al., 2014). If backward masking truly disrupts recurrent processing when associated with a lack of visual awareness (Boehler et al., 2008; Lamme et al., 2002), a feedforward pattern (or one based on local recurrent circuits) is the most likely explanation. Furthermore, the fact that we detect a response to faces, and not to expression, suggests that two different stages of identification and categorization may be supported by qualitatively different mechanisms. It is still the subject of debate whether feedforward processing can support categorization (DiCarlo et al., 2012; Howe, 2017), and our results support the idea that some degree of recurrent processing is necessary (Lamme and Roelfsema, 2000; Maguire and Howe, 2016).

Note that the spatial resolution of MEG prevents us from drawing strong conclusions on the origin of this response to faces. Furthermore, recent observations have been made about concerns of information spreading in source-space MVPA analyses of MEG data, potentially overestimating the spatial extent of effects (Sato et al., 2018). In this chapter, we restricted our source-space decoding analysis to localizing effects identified at the sensor level, and we applied randomization testing with an omnibus threshold in order to avoid spurious effects (3.3.10) and to alleviate the trade-off between maximizing information and reducing false positives.

All expressions presented for at least 30 ms were decodable from MEG data. In Chapter 2, we found early above-chance decoding of angry expressions compared

to happy and neutral faces. In the present chapter, we show early decoding of both face and expression (~100 ms), with only a slight advantage for angry expressions (107 ms; Table 3.3), suggesting a contribution of task-related effects to early visual processing. Furthermore, it is important to note that all analyses described here were performed across facial identity and that stimuli were controlled in terms of low-level properties. The MEG decoding results thus support the idea that expression categorization begins at the early stages of visual perception with rapid processing of emotional cues.

On the other hand, behavioural responses to angry faces were less accurate than those to happy and neutral faces, a finding that stands in contrast to the advantage in decoding angry faces from MEG data found in Chapter 2. It is difficult, however, to directly compare the results of the two chapters, given the different paradigms employed, including different stimulus sets and presentation durations. For example, previous research suggests that angry faces may require longer presentation times to be successfully categorized by participants, compared to happy and neutral faces (Du and Martinez, 2013). The lower performance in categorizing angry faces might also be explained by their variability, as they included both open-mouth and closed-mouth expressions, some of which may have been more difficult to categorize. However, this disadvantage is not reflected in MEG decoding results, which show comparable discriminability of all pairs of expressions based on neural patterns. Furthermore, the evidence for the behavioural effect is not particularly convincing ($P = 0.046$, $\eta^2 = 0.12$). Further research including more extensive angry face sets is needed to assess the generalizability of this finding.

3.5.2 Expression and awareness

In this experiment, we measured subjective visual awareness using a perceptual awareness scale. Subjective and objective measures of awareness both have their strengths and limitations; although subjective measures pose a criterion problem (Szczepanowski and Pessoa, 2007), objective measures (such as performance on a forced-choice task) may reflect unconscious processing (Lau, 2008; Song and Yao, 2016; Wierzchoń et al., 2014). We restricted our experiment to subjective awareness, shown to be effectively captured by perceptual awareness scales, particularly when

employed after discrimination tasks (Sandberg et al., 2010; Wierzchoń et al., 2014). Here, the discrimination task was used to verify subjects' compliance and assess the presence of potential expression biases in responses given to subliminal faces.

It is not surprising that we detected a subliminal response to faces outside of subjective awareness, considering the wealth of evidence on non-conscious face processing (Axelrod et al., 2015). However, in terms of non-conscious expression processing, the results are mixed. Despite the absence of a subliminal expression effect in MEG responses, behavioural data suggest that expression (specifically, a model differentiating between emotional and neutral stimuli) explains approximately one quarter of the variance in behavioural responses given to faces presented for 10 ms. This effect is not revealed by the analysis of individual performance on the task, suggesting that model-based approaches to the analysis of behavioural responses can provide additional information. With the caveat that low numbers of trials were included in this analysis, the fact that cross-subject patterns of response reflected shared variance between the models based on expression, facial features and facial configuration points to a certain degree of expression processing taking place outside of subjective awareness.

The absence of a subliminal expression effect in the neural data may be explained by three main aspects in the study design and analysis: (1) stimuli were normalized in terms of low-level properties, minimizing the detection of visual differences at early stages of perception; (2) we used a cross-identity classification approach, ensuring that we investigate categorical differences; (3) we used a very short stimulus presentation time, reducing the amount of information available to the visual system and limiting the possibility of residual awareness. Although absence of evidence cannot be taken as evidence of absence, we were able to detect a subliminal response to faces despite a lower number of scrambled trials, as well as expression effects to faces presented for longer than 10 ms (using similarly sized datasets). As the MVPA framework and the analysis pipeline were chosen to maximize signal and statistical power, it is likely that this result reflects a true absence of an effect in the MEG data.

3.5.3 Ventral stream representations of behaviour and face configuration

To understand the representations underlying our decoding results, we investigated the similarity between MEG patterns and models based on behavioural performance, as well as facial expression, identity, configuration, and spatial envelope.

We found that ventral stream areas encoded sustained and extensive behavioural representations starting at 120 ms after stimulus onset (Figure 3.10). This suggests that the features extracted in face-responsive cortex are relevant in behavioural decision-making, similarly to evidence found in higher-level object and scene perception (Bankson et al., 2018; Cohen et al., 2017; Groen et al., 2018; Walther et al., 2009) and in line with previous studies showing that the perceptual similarity of faces is represented in neural patterns (Furl et al., 2017; Said et al., 2018).

Moreover, we found representations of face configuration in ventral stream areas, with first-order features being represented earlier and followed by second-order features. Facial configuration has long been thought to play an important part in identity and expression perception (Calder et al., 2000), and in our RSA analysis the configural models show some of the strongest contributions among the nine models tested. In fact, we show that with the exception of a brief time window, no "categorical" representations, as quantified by the high-level models, are formed in occipitotemporal cortex; instead, configural representations appear to overlap with representations of behaviour, suggesting that it is face configuration that drives expression-selective responses in ventral stream areas and guides behaviour. This is also supported by the successful decoding of expression from occipitotemporal cortex.

The contribution of local features prior to the global configuration model adds to evidence suggesting that emotional face perception is supported by the processing of diagnostic features, such as the eyes and mouth (Fox and Damjanovic, 2006; Wegrzyn et al., 2017). Recent studies have shown that the recognition of familiar faces may not rely on holistic face processing, but on specific features (Mohr et al., 2018; Visconti Di Oleggio Castello et al., 2017), and it has been suggested that responses in face-selective areas such as the OFA may represent faces in terms of topological maps or feature-based models (Henriksson et al., 2015). Particularly for

expression perception, feature-based processing provides an efficient mechanism for the rapid extraction of visual cues essential in human interaction, as reflected by the ability of the Action Unit coding system to quantify facial expressions (Ekman and Friesen, 1977; Srinivasan et al., 2016). However, we note that the Action Unit model RDM assessed here did not significantly correlate with the MEG patterns, probably due to the static and brief nature of our stimuli.

Previous studies have shown differential modulation of ERP components by first-order and second-order face configuration. Some studies have shown the P1 and N170 components to encode the former only (e.g. Mercure et al., 2008; Zion-Golumbic and Bentin, 2007), while others have also shown effects of second-order configuration at N170 latencies (Eimer et al., 2011). Furthermore, fMRI studies have reported a division of labour in the face-selective network, with the FFA thought to play a special role in representing both types of configural information (Golarai et al., 2015; Liu and Ioannides, 2010). Recently, it has been suggested that featural and configural processing of even non-face objects elicit face-like responses in the OFA and FFA (Zachariou et al., 2018). Here, we combined the strengths of source-localized MEG data and the RSA framework to tease apart the two models using a single stimulus set. The searchlight RSA analysis revealed that the two models overlap spatially in a right ventral stream area potentially corresponding to the FFA, but are dissociated temporally: for 150 ms faces, representations switch from first-order to second-order at ~300 ms after stimulus onset, bridging previous fMRI and electrophysiological findings.

Furthermore, this two-stage process appears to depend on the amount of information available to the visual system. For 150 ms faces, local and global configuration models make unique, temporally distinct contributions to explaining the data, as shown in the partial correlation analysis. For 30 ms faces, no unique variance is explained by the two models; furthermore, representations are temporally overlapping in the correlation analysis and occur after 300 ms (Figure 3.11). This complements our sensor-level temporal generalization findings: 30 ms faces are processed through a series of transient coding steps at early stages and a stable representation is formed after 300 ms, when both first-order and second-order features are represented. On the other hand, for 150 ms faces, a two-stage process takes place, with

an initial stable representation emerging at M170 latencies and supported mainly by first-order features, and a later representation after 300 ms encoding second-order configuration. Feature representations thus appear to be linked to the late emergence of stable representatons, thought to be reflective of recurrent processing and categorization (Mohsenzadeh et al., 2018b; Tang and Kreiman, 2017). Importantly, this idea is supported by spatially and temporally overlapping behavioural representations in ventral stream areas.

Together, these findings constitute a stepping stone towards a better understanding of high-level representations in face perception. While binary categorical models can estimate high-level representations and task-related processing, the code supporting visual perception is likely to be better understood in terms of behavioural goals and the visual features supporting them. We show that face-responsive cortex dynamically encodes facial configuration starting with first-order features, and that this supports behavioural representations when participants are performing an expression discrimination task. Furthermore, we show that the cascade of processing stages changes with stimulus duration, pointing to the adaptability of the face processing system in achieving goals when visual input is limited. Finally, although we find evidence of a subliminal neural response to faces, we only detect a subliminal response to expression at the behavioural level using a variance partitioning approach. These results bridge findings from previous fMRI and electrophysiological research, revealing the spatiotemporal structure of face representations in human occipitotemporal cortex.

Although they highlight the remarkable adaptability of the visual system in the presence of limited visual input, the findings described in this chapter depend on the explicit processing of expression. In fact, faces are the object of undivided attention both here and in Chapter 2, regardless of the nature of the task. Limiting visual information or presenting participants with an orthogonal task do not address the "automaticity" of expression perception from the perspective of attentional resources: what happens when other stimuli compete for our attention?

Evidence of automatic prioritization of emotional faces suggests that even when

presented as distractors, expression-related cues are processed and impact both behavioural performance and neural responses (Del Zotto and Pegna, 2015; Vuilleumier, 2005). However, whether this is truly automatic or depends on the allocation of attention to competing stimuli is the subject of debate (Chen et al., 2016; Pessoa et al., 2002a; Pessoa, 2005; Pessoa et al., 2002b). In the next chapter, we address this question by presenting emotional faces as distractors in an unrelated task with varying levels of difficulty.

Chapter 4

Emotional face distractors do not capture attention

4.1 Abstract

After evaluating implicit and explicit face perception in previous chapters, the present chapter addresses the processing of emotional face distractors. Previous research suggests that emotional faces are salient enough to be processed even when our attention is engaged elsewhere, but it is still unclear whether this depends on the availability of attentional resources. To address this, we manipulated the difficulty of a grating orientation discrimination task and used a covert spatial attention paradigm to orient attention away from emotional expressions presented as distractors. We investigated expression-related effects in evoked responses, alpha-band activity, and broadband patterns using both univariate and multivariate analyses, but found no evidence of expression processing regardless of task difficulty. This result adds to negative findings that have fueled a longstanding debate, and complements results from Chapter 3 highlighting the importance of task demands in face perception.

4.2 Introduction

In Chapter 3, MEG responses to rapidly presented expressions were not detected when faces were presented outside awareness, despite the presence of a subliminal response to faces. In this final chapter on face perception, we address the related question of whether the salience of emotional faces can affect top-down attention when faces are irrelevant to the task at hand.

The bottom-up capture of attention by emotional faces has been well- documented in behavioural and neuroimaging studies (Carretié, 2014; Mohanty and Sussman, 2013), with much of the evidence supporting a threat advantage hypothesis (Huang et al., 2011; Öhman et al., 2001). To achieve this, the amygdala and orbitofrontal cortex are thought to modulate visual processing at early stages (Lim et al., 2009). The enhanced processing of emotional stimuli observed both in implicit and explicit viewing conditions offers a potential explanation for their salience and its resistance to top-down suppression (Vuilleumier, 2005).

Even when irrelevant or detrimental to the task at hand, emotional faces have been shown to elicit distinct effects, from "popping out" in visual search tasks (Öhman et al., 2001) to interfering with behavioural performance (Hodsoll et al., 2011; Pichon et al., 2012). Expression is thought to interact with attention in an additive or competitive fashion, depending on its role in the task being performed (Feldmann-Wüstefeld et al., 2011; Fenker et al., 2010; Holmes et al., 2005; Huang et al., 2011; Ikeda et al., 2013; Weymar et al., 2011). In spatial attention tasks, faces presented peripherally capture attention (Calvo et al., 2014; Eimer, 2000; Müsch et al., 2016; Stefanics et al., 2012). These results support an automatic view of emotional face perception (Vuilleumier, 2005), whereby expression is processed in the absence of task-related goals, cognitive resources or awareness (Moors and De Houwer, 2006).

However, discrepant findings from behavioural and neuroimaging research point to a more complex interaction between emotion and attention. Studies including emotional faces as distractors during a demanding task have found no expression-specific processing (Chen et al., 2016; Devue and Grimshaw, 2017; Holmes et al., 2003; Koster et al., 2007; Pessoa et al., 2002a,b, 2003; Puls and Rothermund, 2018; Silvert et al., 2007). Other studies show attenuation of affective responses in the

presence of high cognitive load (Morawetz et al., 2010; Pessoa et al., 2005a; Sassi et al., 2014). These results support the idea that the "automaticity" of expression processing depends on cognitive load, consistent with a model postulating limited resources in selective attention (Lavie, 2005).

It has also been suggested that timing dissociates emotional and attentional effects, with emotional salience reflected in an early (pre-attentive) response, and top-down attention reflected in later signals (Inuggi et al., 2014; Liu and Ioannides, 2010; Pourtois et al., 2010); however, such effects have also been explained through insufficient cognitive load (Pessoa, 2010), and given the different tasks and modalities used across studies, it is difficult to support any one conclusion.

Although most findings are not directly comparable, this body of research suggests that many factors may underpin the interplay between attention and emotion: cognitive load and relevance to task may lead to the suppression of emotional stimuli (Oliveira et al., 2013), while individual differences (e.g. in trait anxiety) or face saliency may help override this suppression (Straube et al., 2011).

In this chapter, we investigated the impact of cognitive load on the perception of emotional faces presented as distractors in a covert spatial attention task. This type of task is particularly suited for our question because distractor faces are presented concurrently with target stimuli, and because markers of spatial attention such as alpha desynchronization (Diepen et al., 2016) and the N2pc electrophysiological component (Eimer, 1996) have been well-documented.

Participants viewed bilateral stimulus displays composed of emotional faces and target gratings whose orientation they had to identify. By obtaining individual detection thresholds, we manipulated task difficulty across two blocks. We assessed the impact of emotional distractors on behavioural performance and neural patterns, including evoked responses, broadband signals, and attention-related neural markers. We expected that multivariate methods will help uncover expression-related modulations outside attention that may not be reflected in evoked responses, and that these will vary with cognitive load.

4.3 Materials and Methods

4.3.1 Participants

Twenty-eight healthy volunteers took part in the study (16 female, age range 19-42, mean age 21.78 $\pm$4.7). Written consent was obtained from all participants in accordance with The Declaration of Helsinki, and procedures were approved by the local ethics commitee at the School of Psychology, Cardiff University. Three participants were excluded due to excessive eye movements during the task and all analyses reported were conducted using data from the remaining 25 participants.

4.3.2 Stimuli

The experimental paradigm involved a spatial attention task with gratings as target stimuli and faces as distractor stimuli. Twenty faces with angry, neutral and happy expressions from the NimStim database (Tottenham et al., 2009) were used as distractor stimuli (10 female faces, same stimulus set as in Chapter 3). Face images were pre-processed and matched in terms of low-level properties as in Chapter 3.

Target stimuli were sine wave gratings with a spatial frequency of 4.8 cycles/degree of visual angle and a phase of 1.57 radians. The gratings were equal in size and shape to the face stimuli and were randomly oriented to the left or right by an angle of 60 degrees (for a low difficulty level) or a variable angle individually calculated for each participant (for a high difficulty level). The orientation distribution of gratings appearing contralaterally to each type of emotional face did not significantly differ (proportion of right-oriented gratings for each emotional condition: mean 50% ± 3.6%, F(1.64,39.32)=2.76, P=0.085).

4.3.3 Experimental design

MEG data were recorded while participants performed a grating orientation discrimination task requiring them to correctly identify whether target gratings were tilted to the right or left (Figure 4.1).

Each trial commenced with a centrally presented white fixation cross with a duration pseudorandomly chosen from a uniform distribution between 1.1 and 2 s. A cue then replaced the fixation cross, instructing participants to attend either to

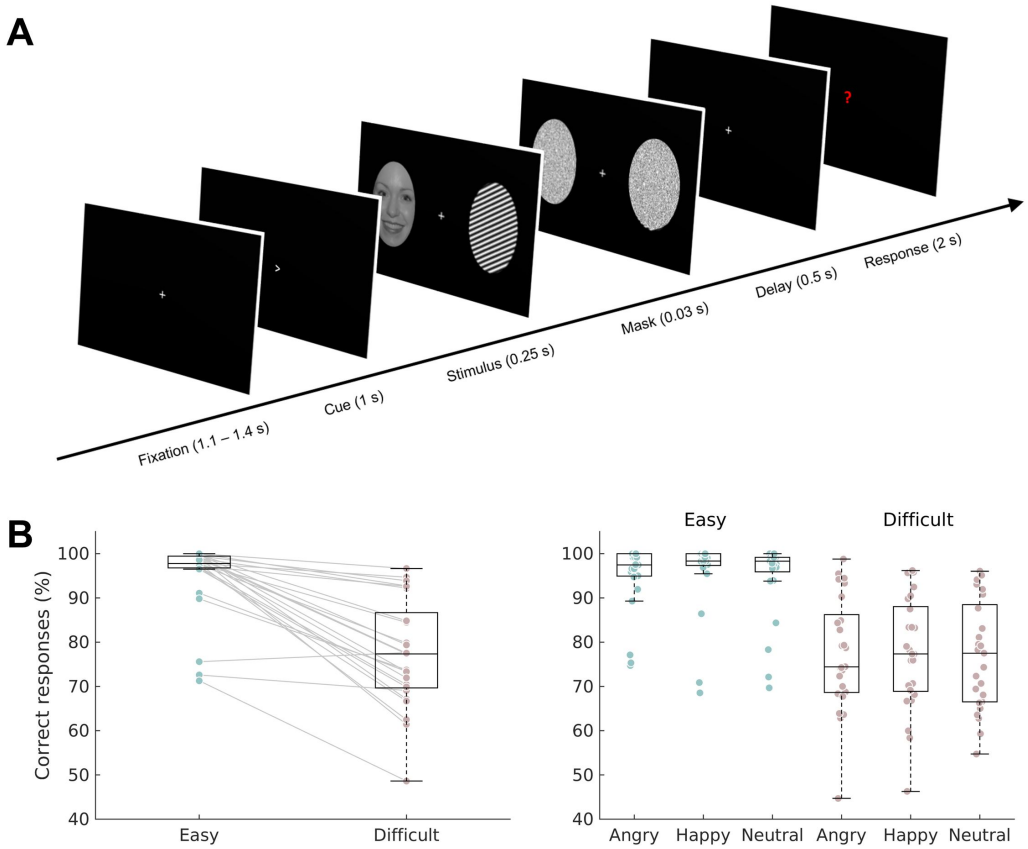

FIGURE 4.1: Overview of the experimental paradigm and behavioural results. **A**. Target stimuli (gratings) and distractors (emotional faces) were presented bilaterally, after a cue indicating the target hemifield. **B**. Participants performed worse in the difficult block (left), but distractor expression did not modulate performance (right). Individual data points are colour-coded according to block difficulty. Boxplots indicate across-participant medians and interquartile ranges.

the left or right hemifield. To avoid working memory effects and lapses in attention and to ensure correct orienting of attention during stimulus presentation, the cue was present on screen until stimulus presentation (Gitelman et al., 1999). The cue duration (1s) ensured predictability of the stimulus, which has been shown to enhance behavioural performance (Nobre, 2001), as well as allowing sufficient time for microsaccades towards the cued location to return to baseline (Engbert and Kliegl, 2003).

Stimulus displays consisted of a grating presented in the cued hemifield and a face distractor presented in the opposite hemifield. The stimuli were presented for 250 ms on a black background approximately 2.04°visual angle to the left and right of the centre of the fixation cross; they were followed by white noise masks shown for ~33 ms in order to prevent aftereffects. The fixation cross remained on screen for 500 ms in order to ensure the dissociation of motor responses from stimulus processing. Participants were then cued by a question mark to make a left/right button press response with their right hand. The paradigm was implemented using Matlab and the Psychophysics Toolbox (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997).

Participants underwent two blocks consisting of two 10-minute sessions each: an easy block (where gratings were always tilted at a 60°angle, with an expected performance close to 100%), and a difficult block (where grating angles were individually determined for each participant, with an expected performance of ~70%).

The order of the blocks was counterbalanced across participants and the difficult block was always preceded by an adaptive staircase procedure performed in the MEG in order to ensure orientation discrimination threshold accuracy (Perry, 2016). The staircase design was similar to the experimental task, but included no faces, and converged on a threshold of 52% correct orientation discrimination using a one-up one-down design with a fixed step size and a ratio of 0.87 between down/up step sizes (García-Pérez, 2001). The staircase started with a grating angle of 3.4°and was constrained to a minimum possible angle of 0.01°. Thirty-five reversals were required for completion of the staircase and the mean of the final 20 reversals was used to determine the discrimination threshold. In order to maintain subjects' attention during the high difficulty block and to ensure that they found it

challenging, but not impossible, we used three different angles in equal proportions during the difficult block: the threshold angle, 80% of the threshold and 120% of the threshold. This ensured that at least one third of the gratings were consistently identifiable, allowing us to minimize learning effects and loss of attention.

4.3.4 Data acquisition

A whole-head structural MRI was acquired for all participants on a General Electric or Siemens 3 Tesla MRI scanner using a 1 mm isotropic Fast Spoiled Gradient-Recalled-Echo pulse sequence in an oblique-axial orientation.

Whole-head MEG recordings were made using a 275-channel CTF axial gradiometer system at a sampling rate of 1200 Hz. Three of the sensors were turned off due to excessive sensor noise. An additional 29 reference channels were recorded for noise rejection purposes and the primary sensors were analysed as synthetic third-order gradiometers (Vrba and Robinson, 2001).

Stimuli were presented on a black background using a ProPixx system with a refresh rate of 120 Hz and a screen resolution of 1920 x 1080 pixels situated at a distance of 1.2 m from the participants. Participants were seated upright while viewing the stimuli and their head position was continuously monitored using electromagnetic coils attached to the nasion and pre-auricular points on the scalp. Participants' head shape was recorded using an ANT Xensor digitizer to aid in co-registration of fiducial locations to the structural MRI scans.

Recordings consisted of four ten-minute blocks (180 trials each) separated by a few minutes' break, with two blocks for each difficulty level. Throughout the experiment, each face image was presented 6 times in each hemifield.

4.3.5 Behavioural data analysis

Behavioural performance was quantified in terms of accuracy (percentage correct trials out of the trials that received a response). Individual accuracies were subjected to a rationalized arcsine transformation (Studebaker, 1985) before being entered into a 2x3 repeated-measures ANOVA with factors *Difficulty* (levels: easy and difficult) and *Expression* (levels: angry, happy, neutral).

4.3.6 Eye gaze data analysis

The participants' right pupil and corneal reflex were tracked using an Eyelink 1000 eye-tracker system with a sampling rate of 1000 Hz. The camera was situated at a distance of 1.2 m in front of the participant. At the start of the experiment, the system was calibrated using a 9-point calibration grid; to account for changes in head position, the eye-tracker was recalibrated after every break.

Vertical and horizontal eye gaze positions were recorded based on pupil position and were analyzed offline using EEGLAB (Delorme and Makeig, 2004), EYE-EEG (Dimigen et al., 2011), and custom Matlab scripts. To assist in rejecting MEG trials, we identified eyetracker trials containing a saccade or fixation to either hemifield during stimulus presentation. To perform statistical analysis, eye gaze data were averaged first within the time window of stimulus presentation, and then across trials and sessions within each difficulty block. Vertical and horizontal eye gaze data were averaged prior to performing a 2 x 3 ANOVA to assess the impact of difficulty and distractor expression. We found no significant effect of task difficulty ($F(1, 23) = 1.5, P = 0.23$), no effect of expression ($F(1.72, 39.67) = 1.13, P = 0.33$) or interaction effect ($F(1.7, 39.13) = 0.33, P = 0.68$).

4.3.7 MEG data preprocessing

MEG data were preprocessed using Matlab and the Fieldtrip toolbox (Oostenveld et al., 2011). Trials containing eye movement and muscle artefacts were rejected after visual inspection; trials containing head motion in excess of 5 mm were also excluded from analysis. We used the eye-tracker data to detect and exclude trials containing fixations or saccades to the stimuli in either hemifield, thus ensuring that only trials where covert attention was truly employed were included in the analysis. This led to a mean of 17.32% ± 14.89% of trials being rejected across participants. Head coil position for each dataset was set to the average across all trials.

To assess the encoding of expression-related information outside attention, we analyzed (1) evoked responses, (2) broadband MEG signals, and (3) alpha-band MEG signals, which have been shown to index covert spatial attention (Kelly et al., 2005). For all analyses, MEG data was preprocessed similarly to methods from

previous multivariate investigations of covert spatial attention (e.g. Gerven et al., 2009; Roijendijk et al., 2013). Trials were low-pass filtered at 100 Hz, de-meaned and downsampled to 300 Hz, with an additional comb filter applied to eliminate the mains noise and its harmonics.

4.3.8 ERF analysis

For evoked response analyses, the data were bandpass-filtered between 0.5 and 30 Hz and axial gradiometer ERFs were converted into planar representations. Differences elicited by stimulus lateralization were assessed across 100 ms time windows using paired t-tests at each sensor and omnibus-corrected randomization testing (5000 iterations). Differences in distractor expression processing were evaluated using repeated-measures ANOVAs with randomization testing separately for each face lateralization condition and difficulty level.

To assess potential effects of distractor expression on markers of spatial attention, we investigated the N2PC component (Eimer, 1996) by calculating responses from right occipital and left occipital axial gradiometers to contralaterally and ipsilaterally presented targets. Responses were averaged separately for each distractor expression, and the ipsilateral average ERFs were subtracted from the contralateral ERF. We then compared this difference wave across expressions using repeated-measures ANOVAs with randomization testing at each 100 ms time window.

4.3.9 Alpha modulation

Alpha-band frequency analysis was performed using a Hanning taper method centred on 10 Hz with a 2 Hz smoothing to effectively obtain a frequency band between 8 and 12 Hz (Bahramisharif et al., 2012). The analysis spanned a time window starting 500 ms after cue onset and ending 800 ms after target onset (1.3 s), minimizing potential eye movement artifacts (Engbert and Kliegl, 2003). Sliding windows of 150 ms with 50 ms overlap ensured that at least one complete oscillatory cycle was included at each frequency. To obtain interpretable spatial patterns, axial gradiometer data were transformed into planar representations (Bastiaansen and Knösche, 2000).

The alpha-band power spectra were averaged over time and used to calculate the sensor-wise alpha modulation (Horschig et al., 2015; Roijendijk et al., 2013) for each participant by contrasting the alpha power for targets presented in the right hemifield (α_R) with the power for targets shown on the left (α_L).

$$\alpha_{mod} = \frac{\alpha_R - \alpha_L}{\alpha_R + \alpha_L} \tag{4.1}$$

The strength of the alpha modulation was statistically assessed for the two difficulty levels using one-sample t-tests against a mean of zero and randomization testing (1000 iterations), with cluster correction for multiple comparisons (cluster-forming $\alpha = 0.05$, cluster $\alpha = 0.025$). Pairwise t-tests were similarly conducted to assess any effects of expression on alpha modulation.

Although we focused on alpha-band activity, previous studies have also shown spatial attention modulations (Koelewijn et al., 2013; Magazzini and Singh, 2017), as well as emotional face distractor effects (Müsch et al., 2016) in the gamma band. To assess this possibility, we performed a similar analysis using a frequency band centered on 70 Hz with 10 Hz smoothing, in order to reproduce the 60-80 Hz frequency band reported in Müsch et al., 2016. Preprocessing, gamma modulation computation, and statistical testing were performed as for the alpha band.

4.3.10 Decoding analyses

Broadband decoding

To assess the differential processing of unattended facial expressions in broadband MEG signals, we performed a time-resolved decoding analysis using anatomically defined sensor sets. Given the nature of the paradigm, the analysis was performed separately for faces presented in the right and left visual field, and for sensor sets in the right and left hemisphere (Figure 4.2). To ensure that informative signals were included, a pooled analysis was also performed combining right and left occipital responses contralateral or ipsilateral to the face stimuli. Binary pairwise decoding of expression was performed as described in Chapter 3 in terms of temporal resolution, trial averaging, multivariate noise normalization, and cross-exemplar five-fold cross-validation.

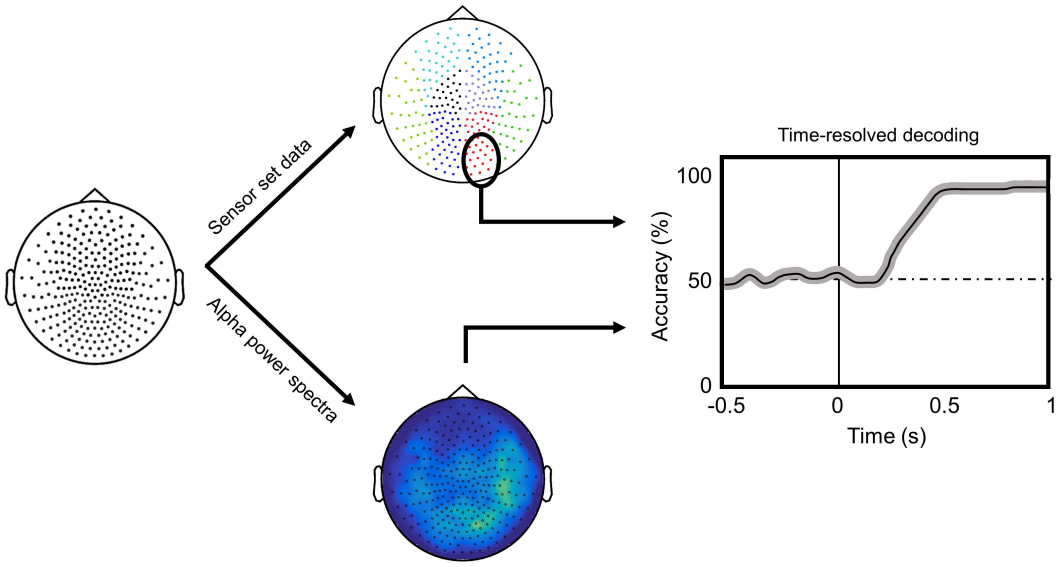

FIGURE 4.2: Overview of the MVPA analysis in this chapter. Classification was performed on 10 anatomically defined sensor sets and on alpha-band power spectra. As no above-chance decoding results were obtained in sensor space, no source-space decoding was performed.

To assess classification performance, accuracies were recomputed using 100 label-shuffling iterations for each participant, decoding problem, difficulty level and stimulus lateralization condition, using the sensor set and time point obtaining the maximum accuracy across subjects (Dima et al., 2018a). P-values were thresholded against the maximal distribution across tests (Nichols and Holmes, 2001; Singh et al., 2003) and a further FDR correction (q=0.05) was applied across time points.

Alpha-band decoding

To assess whether the alpha desynchronization in this experiment was a reliable index of covert attention (Bahramisharif et al., 2012; Tonin et al., 2012; Treder et al., 2011), we also performed multivariate decoding of stimulus laterality based on single-trial power spectra (Figure 4.2): (1) averaged across the analysis window; (2) averaged across the cueing period; (3) time-resolved, using time windows of 150 ms. Classification accuracies were assessed using randomization testing within-subject (following previous research on covert attention decoding, e.g. Bahramisharif et al., 2012).

Next, we performed emotional expression decoding based on alpha-band power spectra to assess potential effects of distractor expression on the strength of alpha desynchronization in our spatial attention task. Expression decoding was performed separately for each time window and evaluated using five-fold cross-validation. Two analyses were performed on feature sets including either all MEG sensors, or sensors showing significant alpha power modulation. Trials were split according to the face hemifield or pooled (Figure 4.5A). To reduce computational cost, 100 label-shuffling iterations were conducted to assess statistical significance; we ensured that the null distribution was conservatively estimated by conducting omnibus thresholding across all tests and setting the alpha level to 0.01 (i.e., no randomized accuracies were allowed to surpass the observed accuracy).

Finally, emotional expression decoding was similarly conducted using the gamma power spectra across the MEG sensor set.

4.3.11 Bayesian statistics

Since most of the above tests revealed no significant expression-related effects, we sought to estimate the strength of the evidence in favour of the null hypothesis by comparing null hypothesis testing results with their Bayesian counterparts. As our original hypothesis entailed an advantage for angry faces in escaping attentional suppression, we focused on the comparison between angry and neutral faces for the purposes of this follow-up analysis. To reduce the number of comparisons performed, we obtained summary measures for each of the signals of interest (grand average ERFs, the N2PC difference wave, and alpha modulation), and compared these across participants using (1) paired t-tests and (2) Bayesian t-tests, implemented in JASP (Version 0.9; https://jasp-stats.org/) using the Summary Stats module (Ly et al., 2018). In all analyses, we used a zero-centered Cauchy distribution with a default scale of 0.707 as the default prior distribution of the population effect size.

The summary measures subject to this analysis were obtained as follows: for evoked response analysis, we averaged responses from (1) occipital, parietal and temporal sensors and (2) sensors found to significantly encode target lateralization, across a time window between 100 and 400 ms (to ensure the capture of any

face-specific responses); for the N2PC component, the difference wave obtained by subtracting the ipsilateral-to-target response from the one contralateral to target was averaged across the 200-400 ms time window (Eimer, 1996); for alpha-band activity, averaging was performed across sensors exhibiting significant alpha modulation.

4.4 Results

4.4.1 Behavioural results

The difference in performance across participants between the easy and difficult tasks suggested that the difficulty manipulation was effective (95% and 77% accuracy respectively; Figure 4.1B). A 2x3 repeated measures ANOVA with factors *Difficulty* and *Expression* on rationalized arcsine-transformed accuracies revealed a significant effect of the difficulty manipulation on performance ($F(1,24) = 85.8, P = 2.14 \times 10^{-}9, \eta^2 = 0.78$). No effect of emotional expression or interaction effect was found ($F(1.72, 41.38) = 0.91, P = 0.4, \eta^2 = 0.04; F(1.97, 47.4) = 0.45, P = 0.64, \eta^2 = 0.02$; Figure 4.1C).

There was no effect of difficulty or emotional expression on eye gaze data averaged across the stimulus presentation duration ($F(1,23) = 1.5, P = 0.23; F(1.72, 39.67) = 1.13, P = 0.33$; and $F(1.7, 39.13) = 0.34, P = 0.68$).

4.4.2 No distractor effects in evoked responses

We found evidence of stimulus lateralization effects (target right vs target left) reflected in evoked responses (Figure 4.3), with significant effects starting at ~150 ms (minimum P=0.0008, maximum t(24) = 5.7). We found no effect of expression across face lateralization conditions and difficulty levels after correction for the number of tests conducted, although there was an effect approaching significance at one left occipital sensor (ML032) for faces presented in the right hemisphere (P=0.037, F(2,48)=10.04, ~225 ms).

To assess potential effects of distractor expression on markers of spatial attention, we also investigated the N2PC component (Eimer, 1996) by calculating responses from right occipital and left occipital MEG sensors to contralaterally and

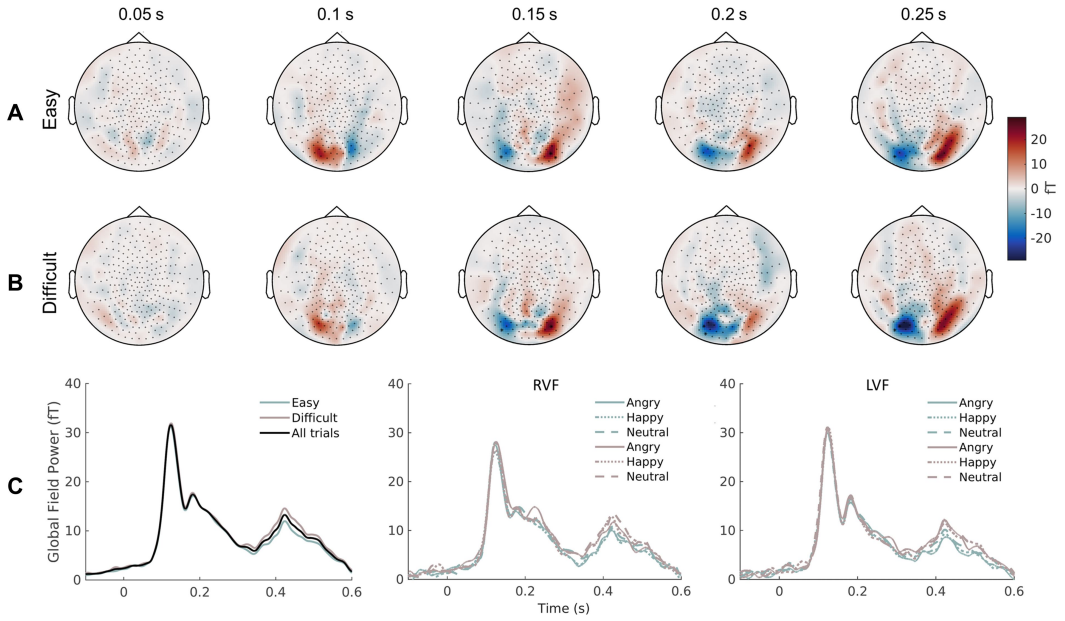

FIGURE 4.3: Evoked response results. **A**. Easy block: grand average difference ERF between trials with a right hemifield target and trials with a left hemifield target. Sensors exhibiting significant differences between right and left hemifield targets are highlighted with asterisks. **B**. As in A for the difficult block. **C**. Global field power across planar gradiometers and subjects, plotted against a 500 ms pre-stimulus baseline for trials from each condition.

ipsilaterally presented targets (Figure 4.4). No significant effect of expression was found across the two difficulty levels (P>0.09, F(2,48)<2.55).

4.4.3 Alpha power and stimulus laterality

To assess the effects of spatial attention on alpha activity, we calculated the alpha modulation for each channel (contrasting alpha activity for trials with a target in the right hemifield with those with a target in the left hemifield; Figure 4.5A). During the easy block, alpha modulation reached a maximum of 0.22 across all subjects in a right occipital cluster (P = 0.02), with two clusters obtained during the difficult block (maximum modulation 0.25, minimum P = 0.004).

When decoding target laterality from the average alpha activity across the entire analysis time window during the easy and difficult blocks, we found above-chance classification in 13 and 17 subjects respectively. However, when decoding across the cue period, the success rate was markedly decreased, with above-chance accuracy in 5/6 subjects out of 25 (Figure 4.5D). This points to inconsistent subject-wise responses despite the group-level effect found, as well as to a potential role

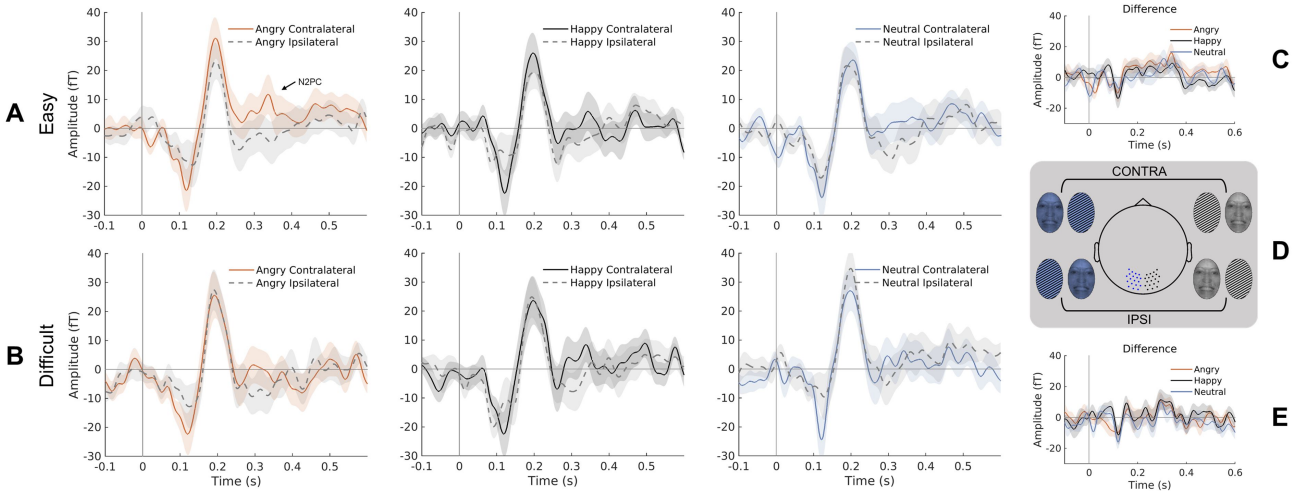

FIGURE 4.4: The N2PC component for each distractor expression. **A,B**. Grand average axial gradiometer ERFs contralateral and ipsilateral to target, separately for each distractor expression. The N2PC occurs at ~200-300 ms. **C, E**. Difference waves between the contralateral and ipsilateral ERFs for each distractor expression during the easy and difficult blocks respectively. **D**. N2PC computation: responses were pooled across the right and left occipital MEG sensor sets for contralateral and ipsilateral targets respectively.

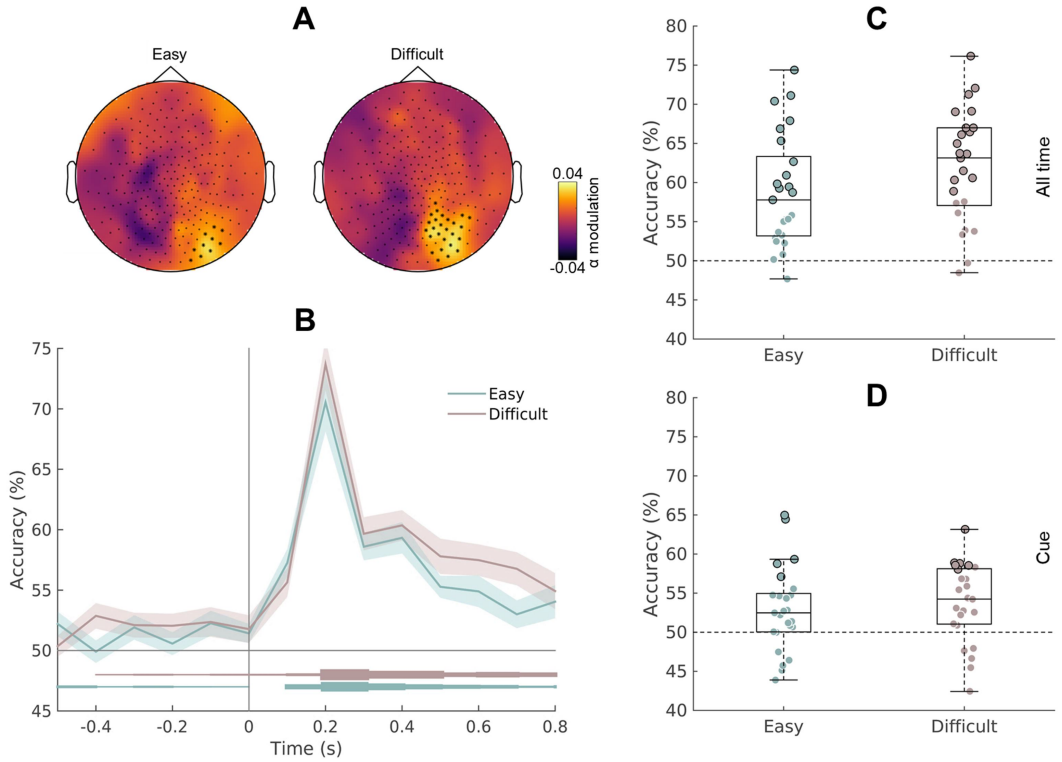

FIGURE 4.5: Alpha modulation. **A**. Sensor maps of alpha modulation (target-right minus target-left) in each block, with significant sensors highlighted (P<0.05, corrected). **B**. Decoding the laterality of stimulus presentation from alpha modulation values. Time-resolved average accuracy traces are shown, with horizontal bars indicating significance (width corresponds to the number of significant subjects). **C**. Decoding the laterality of stimulus presentation from the average alpha modulation values across the decoding time window. Subject-wise accuracies are shown as individual data points, with significant subjects outlined in black. **D** As in C, for the cue time window. Note that fewer subjects achieve above-chance decoding (only 5/6 out of 25).

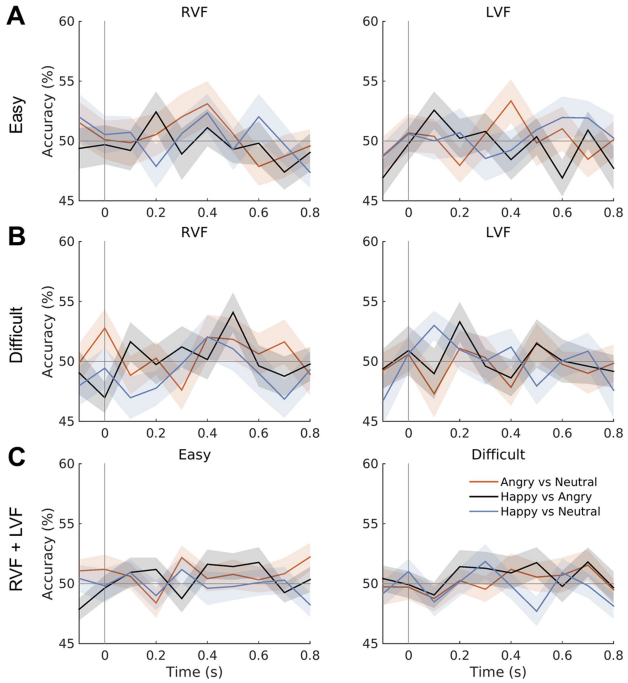

FIGURE 4.6: Alpha-band decoding results using the whole MEG sensor set. **A**. Time-resolved accuracy traces averaged across subjects for faces presented in the right visual field (RVF) or the left (LVF) during the easy block. Shaded areas represent $\pm SEM$. **B**. As in B, for the difficult block. **C**. Decoding results using the pooled dataset (faces presented in both hemifields).

played by visual differences associated with the two lateralization conditions. Indeed, time-resolved decoding using 150 ms time windows shows that only few subjects achieve above-chance accuracy prior to stimulus onset, with a sharp increase in accuracy at ~100 ms and the highest proportion of significant subjects at 200 ms (21 and 24 out of 25 respectively).

Although alpha activity shows the expected laterality effects during our spatial attention task, consistent with results from previous investigations, the study design does not allow us to isolate covert spatial attention during stimulus presentation. However, we may ask whether these stimulus laterality effects are affected by distractor facial expression, irrespective of whether this effect is mediated by attention or visual properties. Expression decoding based on alpha-band power spectra (Figure 4.6) does not rise above chance in any of the subjects, regardless of the hemisphere or difficulty level (maximum accuracy across subjects: 54.09% on the MEG sensor set; 55.4% using feature selection). Thus, while alpha-band activity clearly reflects target laterality, it is not modulated by distractor expression even

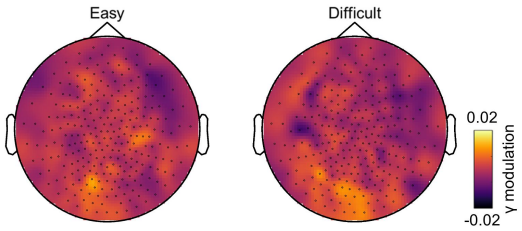

FIGURE 4.7: Sensor maps of gamma modulation (target-right minus target-left) in each block.

when the competing task is not cognitively demanding.

Finally, a similar analysis of gamma-band activity revealed no significant effects of target laterality as reflected in the gamma modulation Figure 4.7, despite a weak lateralized pattern observed in the difficult block. Decoding of emotional expression using the gamma-band power spectra did not achieve accuracies over 54.79% across all tests.

4.4.4 Distractor expression is not decodable from broadband signals

To assess potential differential processing of unattended emotional expressions, we also performed a time-resolved decoding analysis in sensor space (Figure 4.8). We found no above-chance decoding of expression in any of the 10 sensor sets used in the analysis, regardless of the expressions being decoded, the difficulty level, or the face lateralization condition (mean accuracy across subjects, time and tests 50.07% $\pm$ 2.34, range 40.55-58.86%). Pooled decoding analyses of occipital responses contralateral and ipsilateral to the face stimuli also failed to rise above chance level (mean accuracy 50.29% $\pm$ 2.22, range 42.71-58.78%).

4.4.5 Evidence of absence: Bayesian results

We conducted follow-up frequentist and Bayesian paired t-tests on responses to angry and neutral distractors using summary measures of the evoked responses, N2PC, and alpha modulation, in order to quantify the amount of evidence provided by the data. Across 12 tests conducted (Table 4.1), we found moderate or strong evidence for the null hypothesis in 9 tests, and only anecdotal evidence (as labelled in JASP) for the alternative hypothesis in the remaining 3 tests. Note that

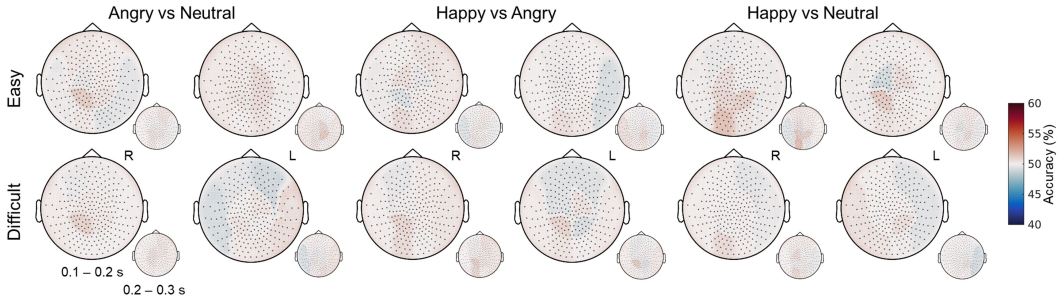

FIGURE 4.8: Expression decoding using anatomically defined sensor sets. Accuracies were averaged using 100 ms time windows and plotted on topographic maps for each decoding problem and face lateralization condition separately. The main plots show results for the 100-200 ms time window, with smaller plots showing similar results for the following time window (200-300 ms). The results shown here are not above the empirically estimated chance level.

in 2 of these latter tests, the effect was in the opposite direction to the one predicted, and all p-values were relatively high ($P > 0.02$ uncorrected).

Furthermore, tests conducted on responses from the difficult blocks tended to provided stronger evidence in support of the null hyposis ($BF_{01} > 4$ in 4 instances), while responses from easy blocks tended to provide less conclusive evidence ($BF_{01} > 4$ in a single test). This could be construed as indirect evidence for the effect of increasing cognitive load in eliminating responses to emotional distractors. Combined with the absence of evidence in our more comprehensive frequentist univariate and multivariate analyses, these results validate the absence of differential MEG responses to distractor expression during this task, especially when task difficulty is increased.

4.5 Discussion

In this chapter, we employed a covert spatial attention task with two levels of difficulty in order to investigate the effects of peripherally presented emotional face distractors. Based on a wealth of evidence on the ability of emotional expressions to capture attention (section 4.2), we expected to find expression-related differences in neural patterns and potentially in behavioural responses. Based on more nuanced models of expression as subject to attentional resource limits (Oliveira et al., 2013), we expected any such effects to decrease or disappear with increasing cognitive

TABLE 4.1: Frequentist and Bayesian t-tests: angry vs neutral distractors

		t(24)	*P*	BF_{10}	BF_{01}	95% CI	Evidence
				ERF: occipital, parietal, temporal			
Easy	Right	-2.28	0.03	1.84	0.54	-0.81, -0.02	H1, anecdotal
	Left	0.76	0.45	0.28	3.62	-0.23, 0.51	H0, moderate
Difficult	Right	-0.38	0.7	0.23	4.43	-0.44, 0.29	H0, strong
	Left	0.13	0.89	0.21	4.71	-0.34, 0.39	H0, strong
				ERF: sensor selection			
Easy	Right	-2.53	0.02	2.9	0.34	-0.88, -0.05	H1, anecdotal
	Left	-0.1	0.92	0.21	4.72	-0.38, 0.35	H0, strong
Difficult	Right	0.54	0.6	0.24	4.16	-0.28, 0.46	H0, moderate
	Left	2.44	0.02	2.45	0.41	0.04, 0.85	H1, anecdotal
				N2PC component			
Easy		0.88	0.38	0.3	3.35	-0.21, 0.53	H0, moderate
Difficult		0.79	0.44	0.28	3.57	-0.23, 0.51	H0, moderate
				Alpha modulation			
Easy		1.17	0.25	-0.39	2.56	-0.16, 0.58	H0, moderate
Difficult		-0.047	0.96	0.21	4.74	-0.38, 0.36	H0, strong

load. Contrary to expectations, we found no robust differences in distractor expression processing, as assessed through a range of different methods.

Although the task was relatively challenging, we found evidence that it operated as expected at all levels: behavioural performance decreased with task difficulty in most subjects (Figure 4.1), eye gaze data did not show any difficulty-related differences, and neural data reflected stimulus lateralization and the expected alpha desynchronization contralateral to target (Figure 4.5).

However, none of these measures were affected by distractor expression, whether analyzed using traditional statistical methods (group ERF analysis) or multivariate methods at the sensor level. (Note that we did not perform source space analyses here, consistent with our approach of using sensor-space decoding as a benchmark for the presence of an effect before exploring its spatial correlates using source-space decoding.)

4.5.1 Threatening stimuli and spatial attention

Attention is thought to help us make sense of the world by suppressing irrelevant information. However, stimuli with high intrinsic saliency are thought to elicit automatic responses, although different models postulate different degrees of automaticity (Vuilleumier and Righart, 2012). Even when processing of emotional stimuli is enhanced during an unrelated task, these effects are not immune from task-related top-down effects, suggesting that rather than bypassing attention, emotion serves as a facilitator. Mounting evidence supports a view of emotional saliency as automatic (in the sense of rapid and involuntary), but subject to suppression by competitive stimuli. Our results support this view: while in Chapter 2 we find an early threat-related response in passive viewing, in the current chapter we find no evidence of expression processing when attention is oriented away from the faces.

Although unexpected, the absence of an effect is not inconsistent with previous research. Investigations using demanding tasks have found no evidence of expression processing outside attention (Chen et al., 2016; Eimer et al., 2003; Koster et al., 2007; Pessoa et al., 2003; Silvert et al., 2007), suggesting that positive results may be driven by a low cognitive load. In the current study, the peripheral and rapid stimulus presentation ensured that even during the easy block, attentional shifts to distractors would be difficult to make without affecting performance on task. While we expected cognitive load to be sufficiently low during the easy block (as reflected in the high performance across participants), other factors, such as motivation and engagement with the task, may have minimized distractor effects.

Some previous studies involving spatial attention tasks have found enhanced processing of emotional unattended faces. However, much of the evidence involves fearful faces (Bishop et al., 2004; Müsch et al., 2016; Pourtois et al., 2006; Stefanics et al., 2012; Vuilleumier et al., 2001), with less consistent evidence for angry faces in cued paradigms (Mohanty et al., 2009; Santesso et al., 2008, but see Ewbank et al., 2009). Although some behavioural studies have found rapid orienting towards or slower disengagement from angry faces (Belopolsky et al., 2011; Calvo et al., 2006), other studies have only found effects in high anxiety individuals (Bradley et al.,

2000; Fox et al., 2008). Thus, evidence on the processing of angry faces outside attention is inconclusive. What is more, much of the positive evidence for unattended expression perception focuses on amygdala responses, which may be specific to fearful faces or more difficult to detect with our current MEG sensor-level analyses. Moreover, many of the experiments reporting expression processing outside attention used tasks in which target features overlapped with irrelevant features (facial expression). A low degree of conjunction between relevant and irrelevant features, as in this chapter, has been shown to degrade irrelevant feature representations (Vaziri-Pashkam and Xu, 2017).

Furthermore, strict normalization of low-level visual features across our stimulus set meant that perceptual differences were less likely to attract attention. It has been suggested in a previous study that peripheral faces attract attention through visual features such as their smiles, rather than affective features (Calvo et al., 2014). Such effects may have been reduced by our stimulus normalization procedure, together with the rapid presentation and masking procedure employed.

Note, however, that the face set used in this chapter is identical to the set used in Chapter 3. With a presentation time as brief as 150 ms, we were able to show expression-specific effects starting at ~100 ms when faces were the object of a task. Here, stimuli were presented for 100 ms longer (albeit peripherally), yet failed to elicit any differential responses when attention was directed towards the opposite hemifield.

A threshold model has been proposed to explain such results (Carretié, 2014), whereby different individual and stimulus-specific factors decide whether an emotional stimulus reaches the required threshold to trigger an exogenous attention effect during a concurrent task. Heterogeneous results from previous studies have started to uncover such potential modulatory factors, but a more systematic evaluation of the conditions necessary for the processing of unattended emotional faces is needed.

Models postulating pre-attentive automatic processing of emotional faces (Pourtois et al., 2010) are also compatible with the current results. Given that the cueing paradigm required attention to be oriented away from the face hemifield prior to stimulus presentation, it can be argued that the required pre-attentive processing

could not take place, as opposed to other paradigms involving face primes (e.g. Müsch et al., 2016) or dot-probe tasks (Santesso et al., 2008). Just as in the case of awareness manipulations (Chapter 3), different spatial attention paradigms can have different effects on stimulus processing. However, the results presented here add to the negative evidence that has cast doubt on the automaticity of expression processing. Our use of a simple perceptual task coupled with concurrently presented, visually matched faces seems to suggest that, when other factors are controlled, expression does not exogenously capture attention.

4.5.2 Limitations and future directions

Some specific aspects of the experimental design employed here make it difficult to draw strong conclusions from these results. Although we find the expected effects due to target lateralization, it is difficult to investigate face processing in the absence of a control condition consisting of scrambled distractors or no distractors; it is possible that a face-related exogenous attention effect takes precedence over any expression-specific processing, as we have found in Chapter 3 in the case of limited awareness. Future studies could explicitly investigate the possibility that face detection is automatic, with the extraction of specific features from faces being influenced by behavioural goals and other factors.

Furthermore, although our analyses of evoked responses, broadband patterns, and alpha-band spectra converge in showing no emotional modulations, expression-related effects may be otherwise represented in the brain or difficult to detect in our current sample. Although group-level analyses included 25 subjects, it is possible that multivariate analyses would have benefitted from larger numbers of trials, as these analyses are performed within-subject. However, such limitations have not precluded successful decoding of expression in previous, similarly designed experiments. Furthermore, Bayesian analysis results suggest that the data reflect some evidence in favour of the null hypothesis, especially during the difficult block, suggesting that this particular task successfully suppressed distractor expression processing.

The results are in line with the conclusions of Chapter 3 concerning the important role played by behavioural goals in shaping face feature representations in

MEG patterns. In future research, manipulating the object and difficulty of the task while keeping stimuli constant might help shed light on the role of endogenous attention in suppressing emotional saliency.

This chapter concludes the part of this thesis dedicated to face perception. Together, the three chapters offer three different perspectives on emotional face perception and converge in pointing out the importance of context and behaviour. In Chapter 2, we saw that passive viewing of emotional faces leads to a threat advantage in terms of neural processing; in Chapter 3, an expression discrimination task elicited early processing of all expressions; and in the present chapter, focus on a concurrent task eliminated any expression-related effects on neural patterns. Together, these results highlight both the "special" nature of face and expression processing, and the adaptability of the visual system in extracting and relaying the most contextually relevant features.

For the final chapter, we turn to a different type of stimuli whose recognition is essential in everyday life, and investigate the extraction of visual features and formation of categorical representations in natural scene perception.

Chapter 5

From features to categories in natural scene perception

5.1 Abstract

Previous chapters discussed how face information is efficiently detected by a highly optimized visual system. This chapter addresses a different, but related question: in navigating our enviroment, how do we efficiently extract information from visual cues? With recent studies painting a complex picture of the neural representations supporting natural scene perception, it is still not well understood how the brain accomplishes the transition between the visual features of our environment and the high-level representations of human cognition. Here, we addressed this using a controlled stimulus set composed of natural scenes from different categories (natural, urban and scrambled) filtered at different spatial frequencies. To investigate the emergence of categorical responses in a task-free setting, we collected MEG data while participants passively viewed the stimuli. Cross-decoding and representational similarity analyses showed that categorical representations emerge in human visual cortex at ~180 ms and are linked to spatial frequency processing. Furthermore, dorsal and ventral stream areas encoded overlapping representations of low and high-level layer activations extracted from a convolutional neural network. These results suggest that neural patterns from extrastriate visual cortex switch from low-level to categorical representations within 200 ms, highlighting the rapid cascade of processing stages essential in human visual perception.

5.2 Introduction

The previous chapters explored the spatiotemporal dynamics of emotional face perception under different tasks using multivariate approaches. This chapter addresses a different domain in visual perception and shows how similar machine learning methods can resolve MEG responses to passively viewed natural scenes. Such stimuli have the advantage of being naturalistic, while exhibiting specific image properties that make them good candidates in disentangling the contribution of visual properties to neural patterns. Furthermore, computational approaches have been applied more often to responses to natural scenes than faces, leading to a growing understanding of the complex sequence of processing stages enabling scene categorization. In this chapter, we investigate featural and categorical representations of scenes in a passive viewing paradigm, and we combine representational similarity analysis with predictions from a feedforward convolutional neural network in order to test the hierarchy of these representations.

Classic models of natural vision predict a succession of stages transforming low-level properties into categorical representations (VanRullen and Thorpe, 2001; Yamins and DiCarlo, 2016). During natural scene perception, the primary visual cortex processes low-level stimulus properties, while extrastriate and scene-selective areas are associated with mid-level and high-level properties. Categorical, invariant representations of scene category are considered the final stage of abstraction (Felleman and Van Essen, 1991; Ungerleider and Haxby, 1994). Scene-selective brain regions such as the parahippocampal place area (PPA), the retrosplenial cortex (RSC), and the occipital place area (OPA) are often thought to represent such categories (Walther et al., 2009) and have been found to respond to high-level stimuli in controlled experiments (Schindler and Bartels, 2016; Walther et al., 2011).

However, this model has been challenged by evidence of low- and mid-level features being processed in scene-selective areas (Kauffmann et al., 2015b; Kravitz et al., 2011; Nasr and Tootell, 2012; Nasr et al., 2014; Rajimehr et al., 2011; Watson et al., 2014; Watson et al., 2016). Studies of temporal dynamics have found overlapping signatures of low-level and high-level representations (Groen et al., 2013; Harel et al., 2016), suggesting co-occurring and co-localized visual and categorical

processing (Ramkumar et al., 2016). Such evidence casts doubt on the hierarchical model and on the usefulness of the distinction between low-level and high-level properties (Groen et al., 2017).

In particular, spatial frequency is thought to play an important part in natural scene perception, with low spatial frequencies mediating an initial rapid parsing of visual features in a "coarse-to-fine" sequence (Kauffmann et al., 2015a). Its role in the processing speed of different features, as well as evidence of its contribution to neural responses in scene-selective areas (Rajimehr et al., 2011), makes spatial frequency a particularly suitable candidate feature for teasing apart the temporal dynamics of low and high-level natural scene processing.

Recent neuroimaging studies of scene perception have used multivariate pattern analysis (MVPA) to highlight the links between low-level processing and behavioural goals (Ramkumar et al., 2016; Watson et al., 2014). In particular, Ramkumar et al. (2016) showed successful decoding of scene gist from MEG data and linked decoding performance to spatial envelope properties, as well as behaviour in a categorization task.

Here, we aimed to dissociate the role of low-level and high-level properties in natural scene perception, in the absence of behavioural goals that may influence visual processing (Groen et al., 2017). In order to do so, we recorded MEG data while participants passively viewed a controlled stimulus set composed of scenes and scrambled stimuli filtered at different spatial frequencies. Thus, we were able to contrast responses to scenes with responses to matched control stimuli, as well as to assess the presence of a categorical response to scenes invariant to spatial frequency manipulations.

Similarly to previous chapters, we used multivariate pattern analysis and representational similarity analysis to explore representations of scene category in space and time and to assess their relationship to low-level properties. We successfully decoded scene category from MEG responses in the absence of an explicit categorization task, and a cross-frequency decoding analysis suggested that this effect is driven by low spatial frequency features at ~170 ms post-stimulus onset. We also show that categorical representations arise in extrastriate visual cortex within

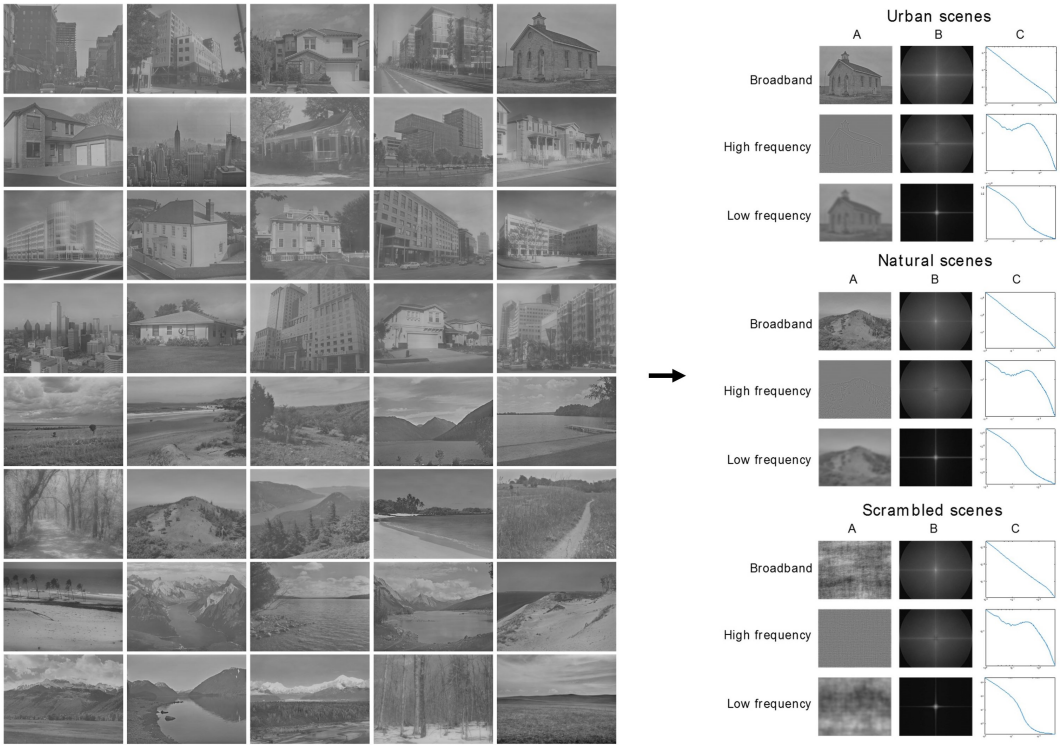

FIGURE 5.1: The complete scene set used in the experiment (**left**), together with examples of filtered stimuli from each condition (**right**, **A**). **B** and **C** show average Fourier and frequency spectra for each condition.

200 ms, while at the same time representations in posterior cingulate cortex correlate with the high-level layers of a deep convolutional neural network (CNN). Together, our results suggest that scene perception relies on low spatial frequency features to create a categorical representation in visual cortex.

5.3 Methods

5.3.1 Participants

Nineteen participants took part in the MEG experiment (10 females, mean age 27, SD 4.8), and fourteen in a control behavioural experiment (13 females, mean age 26, SD 4.4). All participants were healthy, right-handed and had normal or corrected-to-normal vision (based on self-report). Written consent was obtained in accordance with The Code of Ethics of the World Medical Association (Declaration of Helsinki). All procedures were approved by the ethics committee of the School of Psychology, Cardiff University.

5.3.2 Stimuli

Stimuli (Figure 5.1) were 20 natural scenes (fields, mountains, forests, lakes and seascapes) and 20 urban scenes (office buildings, houses, city skylines and street views) from the SUN database (Xiao et al., 2010). Stimuli were 800 $\times$ 600 pixels in size, subtending 8.6 $\times$ 6.4 degrees of visual angle.

All the images were converted to grayscale. Using the SHINE toolbox (Willenbockel et al., 2010), luminance and contrast were normalized to the mean luminance and SD of the image set. Spatial frequency was matched across stimuli by equating the rotational average of the Fourier amplitude spectra (the energy at each spatial frequency across orientations).

To assess the similarity of image amplitude spectra between categories, we calculated pairwise Pearson's correlation coefficients based on pixel intensity values between all images (mean correlation coefficient 0.14, SD 0.27, minimum-maximum range 1.33). Next, we performed an equivalence test (two one-sided tests; Lakens, 2017) in order to compare within-category correlation coefficients from both conditions (i.e., pairwise correlation coefficients between each image and each of the 19 images belonging to the same category) to between-category correlation coefficients (i.e., pairwise correlation coeffients between each image and each of the 20 images belonging to the other category). We assumed correlation coefficients to be similar if the difference between them fell within the [-0.1, 0.1] equivalence interval (Cohen, 1992). Within-category and between-category correlation coefficients were found to be equivalent ($P_1 = 5.3 \times 10^{-11}$, $P_2 = 2.4 \times 10^{-4}$, 90% CI [-0.0025, 0.063]).

To obtain low spatial frequency (LSF) and high spatial frequency (HSF) stimuli, we applied a low-pass Gaussian filter with a cutoff frequency of 3 cycles per degree (25.8 cycles per image) and a high-pass filter with a cutoff of 6 cycles per degree (51.6 cycles per image). Root mean square (RMS) contrast (standard deviation of pixel intensities divided by their mean) was only normalized within and not across spatial frequency conditions, in order to maintain the characteristic contrast distribution typical of natural scenes, which has been shown to influence responses to spatial frequency in the visual system (Field, 1987; Kauffmann et al., 2015a,b).

To produce control stimuli, we scrambled the phase of the images in the Fourier

domain, ensuring equivalent Fourier amplitude spectra across the original and scrambled images (Perry and Singh, 2014). For each spatial frequency condition, we randomly selected 10 of the 20 phase-scrambled images for use in the experiment in order to maintain an equal number of stimuli across conditions (natural, urban and scrambled). The final stimulus set contained 180 images (filtered and unfiltered scenes and scrambled stimuli; Figure 5.1).

5.3.3 Behavioural experiment

Design and data collection

To assess potential differences in the recognizability of different scenes, participants in the behavioural experiment viewed the stimuli and were asked to categorize them as fast as possible. The design of the behavioural experiment was similar to the MEG experiment, but included a practice phase (10 trials) before each block. Participants underwent two blocks in which they had to judge whether stimuli were scenes or scrambled stimuli, or whether scene stimuli were natural or urban respectively. Blocks were separated by a few minutes' break and their order was counterbalanced across subjects.

Images were presented on an ASUS VG248QE LCD monitor with a resolution of 1920 x 1080 pixels and a refresh rate of 60 Hz. Participants were required to make a keyboard response (using the keys *J* and *K*, whose meanings were counterbalanced across subjects), as soon as each image appeared on screen. We recorded responses and reaction times using Matlab and the Psychophysics Toolbox.

Data analysis

To assess the effect of spatial frequency filtering on performance in the categorization task, one-way repeated-measures ANOVAs were performed on individual accuracies (after performing a rationalized arcsine transformation; Studebaker, 1985) and on mean log-transformed reaction times for each categorization task (four tests with a Bonferroni-adjusted $\alpha = 0.0125$). Significant effects were followed up with post-hoc Bonferroni-corrected paired t-tests.

5.3.4 MEG data acquisition

For source reconstruction purposes, in all participants, we acquired whole-head structural MRI scans on a General Electric 3 T MRI scanner using a 1 mm isotropic Fast Spoiled Gradient-Recalled-Echo pulse sequence in an oblique-axial orientation.

Whole-head MEG recordings were made using a 275-channel CTF axial gradiometer system at a sampling rate of 1200 Hz. Three of the sensors were turned off due to excessive sensor noise. An additional 29 reference channels were recorded for noise rejection purposes; this allowed the primary sensors to be analysed as synthetic third-order gradiometers using a linear combination of the weighted reference sensors (Vrba and Robinson, 2001).

Stimuli were centrally presented on a grey background using a gamma-corrected Mitsubishu Diamond Pro 2070 CRT monitor with a refresh rate of 100 Hz and a screen resolution of 1024 $\times$ 768 pixels situated at a distance of 2.1 m from the participants. There were 9 conditions (natural scenes, urban scenes and scrambled scenes filtered at low frequency, high frequency or unfiltered). Each image was presented 4 times, amounting to 80 trials per condition. Participants underwent two recording sessions separated by a few minutes' break.

The data were collected in 2.5 s epochs centred around the stimulus onset. Stimuli were presented on screen for 1 s and were followed by a fixation cross for a varying ISI chosen pseudorandomly from a uniform distribution between 0.6 and 0.9 s. Participants were instructed to press a button whenever the fixation cross changed colour during the ISI. The paradigm was implemented using Matlab and the Psychophysics Toolbox and was adapted from the experimental paradigm described in Chapter 1.

Participants were seated upright during the experiment and electromagnetic coils attached to the nasion and pre-auricular points on the scalp were used to continuously monitor their head position. For co-registration with the structural MRI scans, high-resolution digital photographs of the coil positions were acquired.

5.3.5 MEG analyses

The data were pre-processed using Matlab and the FieldTrip toolbox. Trials containing excessive eye or muscle-related artefacts were excluded based on visual inspection. Condition information was not available during artefact rejection, and there was no significant difference in the proportion of trials rejected between conditions (F(1.5,27.09)=3.33,P= 0.063, 3 $\times$ 3 ANOVA). To account for head motion, we excluded trials with maximum motion of any individual fiducial coil in excess of 5 mm. To account for potential changes in the participants' head position over time, head coil position relative to the dewar was changed to the average position across all trials. Prior to all analyses, the data were downsampled to 600 Hz, baseline corrected using a time window of 500 ms prior to stimulus onset, and a 50 Hz comb filter was used to remove the mains noise and its harmonics.

To test for scene-selective responses present in the event-related fields (ERFs), MEG data were bandpass-filtered between 0.5 and 30 Hz. Axial gradiometer ERFs were realigned to a common sensor position (Knösche, 2002) and averaged across subjects. Based on local minima in the global field power across all trials (Figure 5A), we identified three time windows of interest (Perry and Singh, 2014): 84-143 ms, 143-343 ms, and 343-401 ms. For each time window, we tested for differences between responses to unfiltered (broadband) scenes and scrambled stimuli at all MEG sensors, using paired t-tests and randomization testing (5000 iterations, corrected for multiple comparisons using the maximal statistic distribution).

Prior to sensor-space MVPA analyses, the data were bandpass-filtered between 0.5 and 100 Hz. To test for differences between conditions present in single trials, a linear L1 Support Vector Machine (SVM) classifier was applied to sensor-level data. The classifier was implemented in Matlab using the Statistics and Machine Learning Toolbox and the Bioinformatics Toolbox.

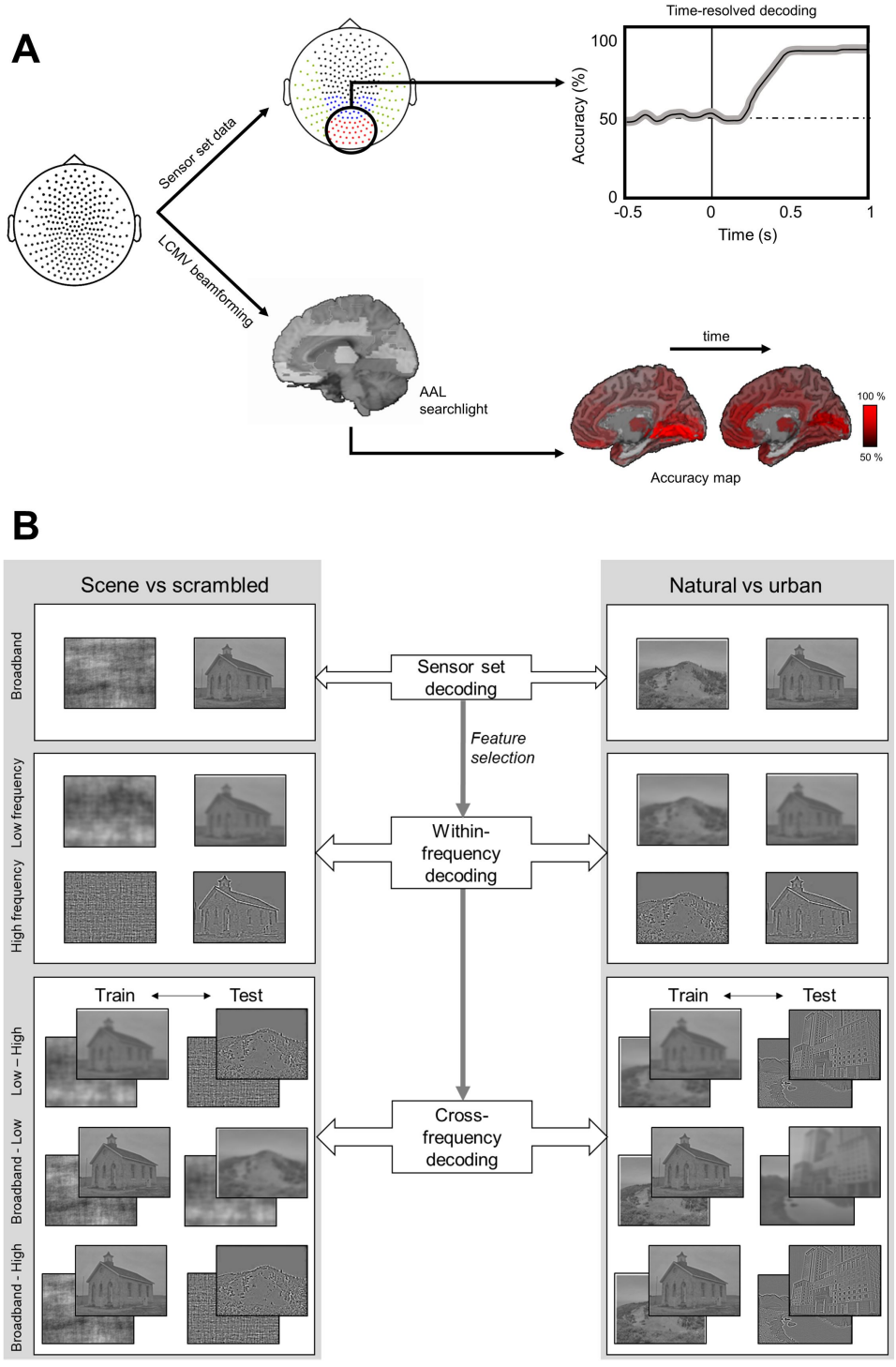

FIGURE 5.2: MPA analysis framework used in this chapter. **A**. Time-resolved decoding was performed on (1) sensor-level data from four anatomical subsets, and (2) source-space data, using an anatomically informed searchlight approach. **B**. Sensor-space analysis pipeline in terms of the stimulus sets used in decoding. Note that in cross-decoding each stimulus set acted in turns as a training and test set, with resulting accuracies averaged across the two cases. Cross-exemplar five-fold cross-validation was performed for all analyses.

5.3.6 Decoding responses to unfiltered scenes

Sensor-space MVPA

A first MVPA analysis (Figure 5.2) was performed on responses to unfiltered stimuli using single-trial data from four anatomically defined sensor sets (occipital, temporal, parietal and fronto-central). Binary time-resolved classification was applied to broadband scenes and scrambled stimuli, as well as broadband natural and urban scenes. As the former problem entailed unequal class sizes, majority class trials were randomly sub-sampled.

The classifier was applied to each time point between 0.5 s pre-stimulus onset and 1 s post-stimulus onset after resampling the data to 600 Hz, thus giving a temporal resolution of ~1.6 ms. Feature vectors were standardized using the mean and standard deviation of the training set. To evaluate classifier performance within subjects, we used cross-exemplar five-fold cross-validation, whereby the classifier was iteratively trained on trials corresponding to 16 of the 20 stimuli from each condition and tested on the remaining 4 stimuli. This ensured that classification performance was not driven by responses to particular visual features repeated across the training and test sets, whilst achieving balanced training and test sets and reducing variability in classification performance.

Source-space MVPA

To perform classification in source space, data in all trials regardless of condition were bandpass-filtered between 0.5 and 100 Hz. We used the FSL Brain Extraction Tool (Smith, 2002) to extract the brain surface from the participants' structural MRI scans and we projected the data into source space using the LCMV beamformer (Van Veen et al., 1997). The forward model (a single-shell sphere) was combined with the data covariance matrix (Hillebrand et al., 2005) to obtain the spatial filter. We defined the source space using a template grid with a resolution of 10 mm that was warped to each participant's MRI in order to ensure equivalence of sources across participants. For each voxel, we independently derived the output as a weighted sum of all MEG sensor signals.

The decoding analysis was performed using an anatomically informed searchlight approach based on the AAL atlas (Tzourio-Mazoyer et al., 2002). For each subject, time-resolved classification with cross-exemplar cross-validation as described above was performed iteratively using the timecourses of sources from each AAL region of interest (ROI), excluding the cerebellum and some deep structures. We chose this approach to reduce computational cost, to improve interpretability across studies and modalities (Hillebrand et al., 2012), and to overcome some of the caveats of traditional searchlight analyses, which assume that information is uniformly distributed in the brain (Etzel et al., 2013).

5.3.7 Using MVPA to evaluate the role of spatial frequency

To maximize the amount of informative features input to the classifier, we performed the next MVPA analyses using the occipital sensor set, which achieved the best classification performance in the broadband scene vs scrambled decoding problem. This ensured minimal overlap between the decoding problem used in feature selection and the follow-up analyses (Figure 5.2).

Decoding responses to filtered stimuli

Despite the use of matched control stimuli, successful decoding of unfiltered scenes does not allow us to disentangle low-level and high-level responses, as differences in local low-level properties cannot be ruled out. Thus, to assess the role played by spatial frequency, we performed scene category decoding (scenes vs scrambled stimuli and natural vs urban scenes) within each spatial frequency condition (HSF and LSF) using the occipital sensor set and cross-exemplar cross-validation.

Cross-decoding

Next, we aimed to test whether scene category representations generalized across spatial frequency categories. To this aim, we trained and tested sensor-space scene category classifiers across different spatial frequency conditions. The analysis was repeated for all three condition pairs using five-fold cross-exemplar cross-validation,

with each set of stimuli acting as a training set and as a test set in turns and the final accuracy averaged across the two cases (Figure 5.2B).

In this analysis, classifier performance was interpreted as an index of the similarity of scene-specific responses across spatial frequency manipulations. Successful decoding across LSF and HSF stimuli would indicate a truly spatial frequency-independent categorical distinction, as there are no overlapping spatial frequencies across the two sets. On the other hand, cross-decoding across unfiltered and LSF or HSF scenes would allow us to detect any spatial frequency preference in the encoding of scene-specific information.

The fact that RMS contrast was not normalized across spatial frequency conditions introduced a potential confound in this analysis. This was not an issue when training and testing within one spatial frequency condition (as RMS contrast was normalized across stimulus categories within each spatial frequency condition). However, both local and global amplitude characteristics were similar between broadband and LSF scenes due to the $1/f$ amplitude spectrum of natural scenes discussed above; this posed a specific concern to the cross-decoding of broadband and LSF scenes. This issue was addressed by conducting cross-exemplar cross-validation. Normalization of low-level features within training and test sets ensured that global contrast characteristics would not be exploited in classification, while testing on novel exemplars ensured that the classifier would not simply "recognize" local features (including contrast) unaffected by the spatial frequency manipulation. This does not preclude the existence of local characteristics that distinguish scenes from scrambled stimuli; however, such characteristics can be expected to be informative in the emergence of a high-level response.

Significance testing

Averaged accuracy across subjects (proportion correctly classified trials) was used to quantify decoding performance, and the significance of classifier accuracy was assessed through randomization testing (Nichols and Holmes, 2001; Noirhomme et al., 2014). To this end, 1000 randomization iterations were performed for each subject, whereby class labels were shuffled across the training and test sets before recomputing classification accuracy. The null distribution was estimated based on

the time point achieving maximum overall accuracy in the MVPA analysis. For time-resolved sensor-space decoding analyses, P-values ($\alpha = 0.01$) were omnibus-corrected using the maximum accuracy across all tests performed (Nichols and Holmes, 2001; Singh et al., 2003), and cluster-corrected across time. To determine 95% confidence intervals around decoding onset latencies, individual decoding accuracies were bootstrapped 1000 times with replacement, and differences in onset latencies were tested using a Wilcoxon signed-rank test. For searchlight decoding in sensor and source space, P-values ($\alpha = 0.001$) were thresholded using the maximum accuracy across sensor clusters/ROIs and cluster-corrected across time.

5.3.8 Representational Similarity Analysis (RSA)

In order to evaluate low and high-level representations of stimuli in our data, we assessed correlations between representational dissimilarity matrices (RDMs) based on temporally and spatially resolved MEG patterns and two sets of models: (1) explicit feature-based models (based on either stimulus properties or stimulus categories), and (2) models extracted from the layers of a deep CNN (Figure 5.3). The second analysis was performed to assess whether evaluating an explicitly hierarchical set of models would support our initial conclusions.

Feature-based models

In order to assess the contributions of low-level features and categorical distinctions, we evaluated four model RDMs based on stimulus properties (Figure 5.3). Visual features were assessed using two models: a low-level model based on spatial frequency, and a mid-level model reflecting the spatial envelope of the images. The former was based on pairwise Euclidean distances between the spatial frequency spectra of the images; the latter was computed using the GIST descriptor (Oliva and Torralba, 2001), which applies a series of Gabor filters at different orientations and positions in order to extract 512 values for each image. These values represent the average orientation energy at each spatial frequency and position and were used to compute pairwise Euclidean distances.

For high-level representations, we used a category-based and an identity-based model. In the former model, all scenes within a category (such as urban scenes)

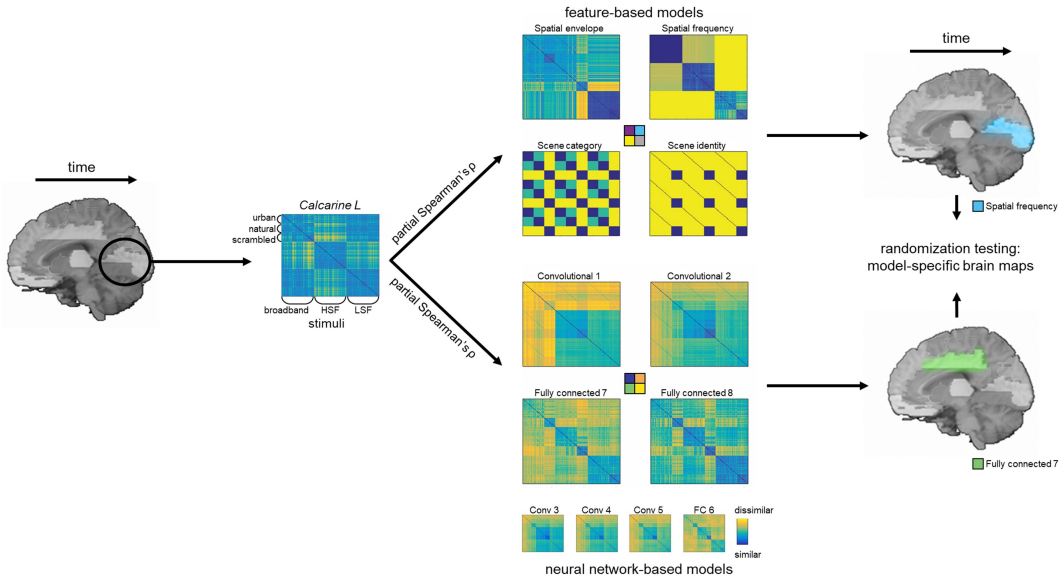

FIGURE 5.3: RSA analysis framework used in this chapter. Time-resolved neural dissimilarity matrices were created for each AAL region and compared to two sets of model dissimilarity matrices, based on either image features or CNN layer activations. Note that 8 models were obtained from CNN layers, but some of these were highly correlated and are only shown as thumbnails (see Figure 5.5). Randomization testing was used to create time-resolved representational brain maps showing the unique contribution of each model to the neural patterns.

were assigned a distance of 0, while scrambled stimuli and scenes were assigned a maximal distance of 1, and distances between different categories of scenes (natural and urban) were set to 0.5. The scene identity model assigned dissimilarity values of 1 to all pairs of natural scenes regardless of category (while all scrambled stimuli were deemed maximally similar). For both models, these values were constant across spatial frequency manipulations.

CNN-based models

To more directly assess the hierarchical processing of our stimulus set in the visual system, we tested a second set of models based on the layers of a feedforward CNN. Using Matlab and the Neural Network Toolbox, we extracted features from an eight-layer CNN pre-trained using the Caffe framework (Jia et al., 2014) on the Places database, which consists of 2.5 million images from 205 scene categories (Zhou et al., 2014). The neural network was a well-established AlexNet CNN (Krizhevsky et al., 2012) with five convolutional layers and three fully-connected

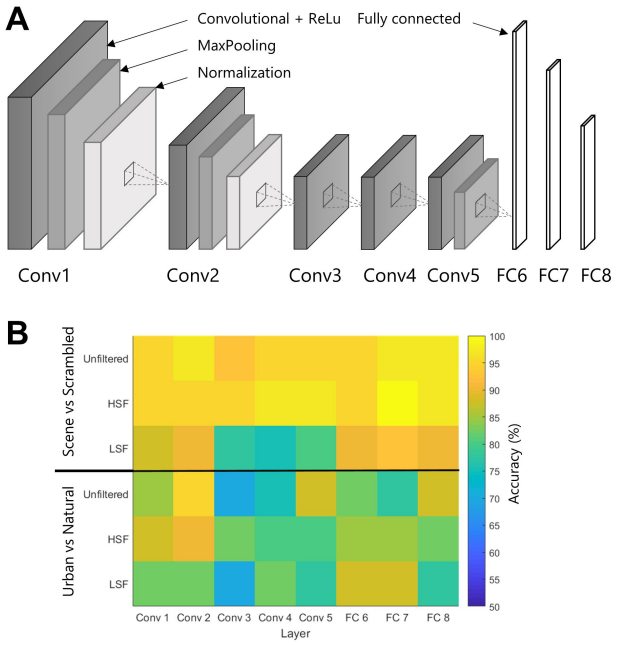

FIGURE 5.4: Convolutional neural network architecture and performance. **A**. The CNN architecture used for model RDM generation. **B**. Accuracy obtained using features from each of the 8 CNN layers for the two decoding problems (5-fold cross-validation). Conv: convolutional; FC: fully connected.

layers (Figure 5.4A). This network architecture has been shown to perform well in explaining object and scene representations in the visual system (e.g. Cichy et al., 2016; Rajaei et al., 2018). We extracted network activations from the last stage of each CNN layer for each image in our stimulus set, and we calculated pairwise Euclidean distances between the feature vectors to obtain eight CNN-based RDMs (Figure 5.3). To assess how well scene categories were represented by these features, we also performed cross-validated binary classification (scene vs scrambled and urban vs natural images) using layer activations, and found high decoding accuracies in all layers (>70%; Figure 5.4B).

RSA analysis framework

In order to assess correlations between model RDMs and neural patterns, MEG data were pre-processed and projected into source space as described above. Neural patterns were computed using source timecourses within each AAL-based ROI for each 16 ms time window after stimulus onset in order to decrease computational cost. Responses to repeated stimuli were averaged within and across subjects and

the Euclidean distance between each pair of stimuli was computed to create neural RDMs.

For each ROI and time window, we computed Spearman's rank partial correlation coefficients between the neural dissimilarity matrix and each of the feature-based models and CNN-based models (Nili et al., 2014). This allowed us to quantify the unique contribution of each model, while controlling for correlations between models. In order to evaluate the impact of RMS contrast on both low-level and high-level category processing, the feature-based analysis was repeated with the RMS contrast-based RDM partialled out. For the purposes of this analysis, RMS contrast was defined as the standard deviation of pixel intensity values divided by mean intensity across each image (Scholte et al., 2009), and the contrast-based RDM consisted of pairwise Euclidean distances between stimulus RMS contrast values (Figure 5.12B).

The significance of the correlation coefficients was assessed through randomization testing, by shuffling the stimulus labels and recomputing the partial correlations 100 times for each ROI and time window. We used a one-sided test, as negative correlations between distance matrices were not expected and would be difficult to interpret (Furl et al., 2017). P-values obtained were thresholded using the maximum correlation coefficient across time points and the alpha was set to 0.01 to account for the number of models tested. This method only highlighted correlations that were stronger than all those in the empirical null distribution.

To assess the maximum possible correlation given the noise in the data, we used guidelines suggested by Nili et al. (2014). We computed an upper bound of the noise ceiling by correlating the average neural RDM across subjects to each individual's neural RDM for each ROI and time window (overfitting and thus overestimating the true model correlation), and a lower bound by correlating each individual's RDM to the average of the remaining 18 subjects' RDMs (underfitting and thus underestimating the correlation).

5.3.9 Eye gaze data collection and analysis

An SMI iView X eyetracker system (SensoMotoric Instruments) with a sampling rate of 250 Hz was used to track the subjects' right pupil and corneal reflection

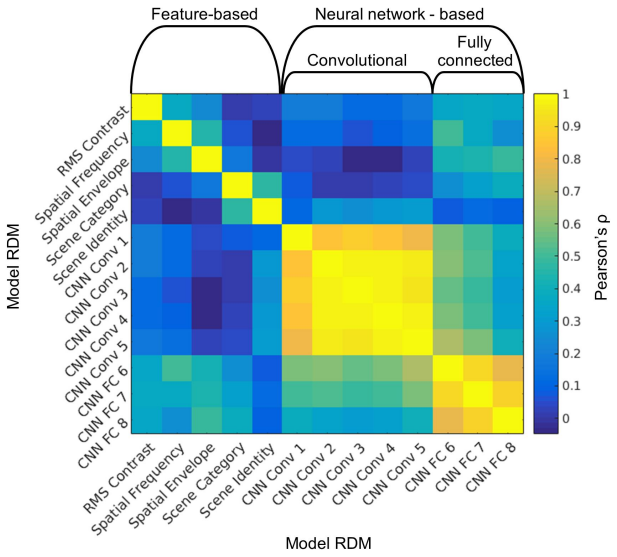

FIGURE 5.5: Correlations between all model RDMs. RDMs based on convolutional layers and fully connected layers of the CNN are highly correlated.

during the MEG recordings. The camera was located in front of the participant at a distance of 120 cm. The system was calibrated using a 9-point calibration grid at the start of each session, and was recalibrated between sessions to account for changes in head position during the break.

Eye-tracker data was analyzed using Matlab, EEGLAB (Delorme and Makeig, 2004), and EYE-EEG (Dimigen et al., 2011). Vertical and horizontal eye gaze positions were recorded based on pupil position and were compared offline in order to assess differences between eye movement patterns across scene categories. After selecting time windows corresponding to the stimulus presentation (1 s post-stimulus onset), portions of missing eye-tracker data corresponding to blinks were reconstructed using linear interpolation prior to statistical analysis. Trials deviating from the mean by more than 2 standard deviations were excluded. We calculated the grand means, medians and standard deviations of eye gaze position for each condition and participant and tested for differences using two-way repeated measures ANOVAs with factors *Category* (levels *natural*, *urban*, and *scrambled*) and *Frequency* (levels *LSF*, *broadband*, and *HSF*). P-values were corrected for six comparisons (three tests on horizontal and vertical eye gaze data). No significant differences were found for either of the two factors ($F(2,36) < 2.57, P > 0.09$ (Category); $F(2,36) < 2.32, P > 0.11$ (Frequency); $F(4,72) < 2.55, P > 0.04, \alpha = 0.0083$).

Next, we performed MVPA to test whether scene categories could be differentiated using single-trial eye gaze data. Gaze position values for the entire stimulus duration were entered as features in an initial analysis, while a subsequent analysis used time windows of 40 ms to check for time-resolved effects. Binary classification was performed on all six pairs of scene category conditions (scenes vs scrambled stimuli and natural vs urban scenes, for each spatial frequency condition). Accuracy did not exceed 51.98% (SD 6.08%) across participants for any of the 6 pairs of conditions tested. Time-resolved MVPA led to similar results (maximum accuracy over time and classification problems 53.69%, SD 5.94%).

5.4 Results

5.4.1 Behavioural categorization results

Participants were asked to categorize stimuli as scenes/scrambled and natural/urban respectively. Performance was high on both tasks (mean accuracy 95.27%, SD 5.63%, and 94.46%, SD 3.56% respectively; Figure 5.6) and ranged between 90.47% and 98.45% across all conditions. We evaluated differences in performance and reaction time between spatial frequency conditions using one-way repeated ANOVAs.

Recognition performance did not significantly differ for scenes filtered at different spatial frequencies when participants had to make urban/natural judgements ($F(1.78, 23.09) = 0.15, P = 0.83, \eta^2 = 0.01$). However, a significant difference was found when participants categorized stimuli as scenes or scrambled stimuli ($F(1.47, 19.09) = 15.44, P = 0.0002, \eta^2 = 0.54$), with LSF images categorized significantly less accurately than broadband ($t(13) = 3.08, P = 0.008$, 95% CI [1.17, 24.43]) and HSF images ($t(13) = 6.03, P = 4.24 \times 10^{-5}$, 95% CI [9.48, 25.94]).

Responses were slightly slower on the scene vs scrambled task (mean raw RT 537 ms, SD 54 ms, versus 506 ms, SD 61 ms on the natural vs urban task). A one-way repeated measures ANOVA on mean log-transformed reaction times revealed a significant effect of frequency for the scene vs scrambled task ($F(1.75, 22.77) = 48.62, P = 1.4 \times 10^{-8}, \eta^2 = 0.79$), with Bonferroni-corrected follow-up tests revealing significantly slower reaction times for LSF images compared to both broadband images ($t(13) = 8.37, P = 1.3 \times 10^{-6}$, 95% CI [0.07, 0.15] and HSF images ($t(13) =$

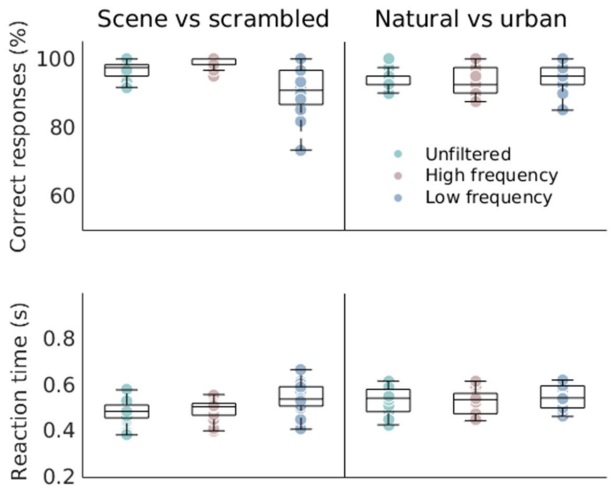

FIGURE 5.6: Categorization performance and mean reaction times for the 14 participants in the behavioural experiment, represented separately for each of the spatial frequency conditions used in the experiment.

$6.92, P = 10^{-5}$, 95% CI [0.05, 0.12]). A smaller effect was found for the natural vs urban task ($F(1.71, 22.25) = 6.11, P = 0.01, \eta^2 = 0.32$), with slower reaction times for LSF than HSF images revealed in follow-up tests ($t(13) = 3.06, P = 0.009$, 95% CI [0.01, 0.06]). Despite the effect reported here, we note that performance was above 90% on all conditions, suggesting high scene recognizability regardless of spatial frequency filtering.

5.5 Evoked responses to scenes

To test for scene-selective responses present in the event-related fields (ERF), we assessed differences between conditions at all MEG sensors in three time windows of interest (84-143 ms, 143-343 ms, and 343-401 ms), using paired t-tests and randomization testing.

The largest amplitudes in response to scenes in this dataset were found over occipital and temporal sensors (Figure 5.7). Significant differences in the response to scenes and scrambled scenes were found over temporal sensors (343-401 ms; $P < 0.01, t(18) > 4.72$). At the P2 latency, differences were present between scenes and scrambled stimuli at two occipital sensors, but they failed to survive correction for multiple comparisons (143-343 ms; $P > 0.034, t(18) < 4.4$). No significant differences between scenes and scrambled scenes were found at the P1 latency

($P > 0.4, t(18) < 2.81$).

5.5.1 Decoding responses to unfiltered scene categories

Sensor-space decoding

To evaluate differences in neural responses between stimulus categories, we performed time-resolved decoding of responses to scenes vs scrambled stimuli and natural vs urban scenes using anatomically defined sensor sets. Above-chance decoding performance was achieved using the occipital sensor set starting at 172 ms and 105 ms post-stimulus onset respectively (Figure 5.8). This effect was transient for both decoding problems; the return to chance level could reflect the absence of late task-related processing in our passive viewing paradigm. There was a significant difference between onset latencies for the two decoding problems ($Z = 26.46, P < 0.001$, 95% CI [13, 97] ms]), likely to reflect earlier decoding of systematic low-level differences between urban and natural stimuli (for example in terms of cardinal orientations). Classification on the parietal sensor set also achieved significance after 318 ms for the scene vs scrambled decoding problem, suggesting more sustained scene processing along the dorsal stream (Table 5.1).

Source-space decoding

To spatially localize the effects revealed by sensor-space MVPA, we moved into source space and performed MVPA analysis of scene category processing using virtual source timecourses obtained through LCMV beamforming and an AAL atlas-based ROI searchlight approach.

Accuracies obtained in source space were comparable to sensor space performance (Table 5.1). Above-chance decoding was achieved for both problems in calcarine cortex (105 and 215 ms respectively) and along the dorsal stream for the scene versus scrambled decoding problem (230 ms; Figure 5.9).

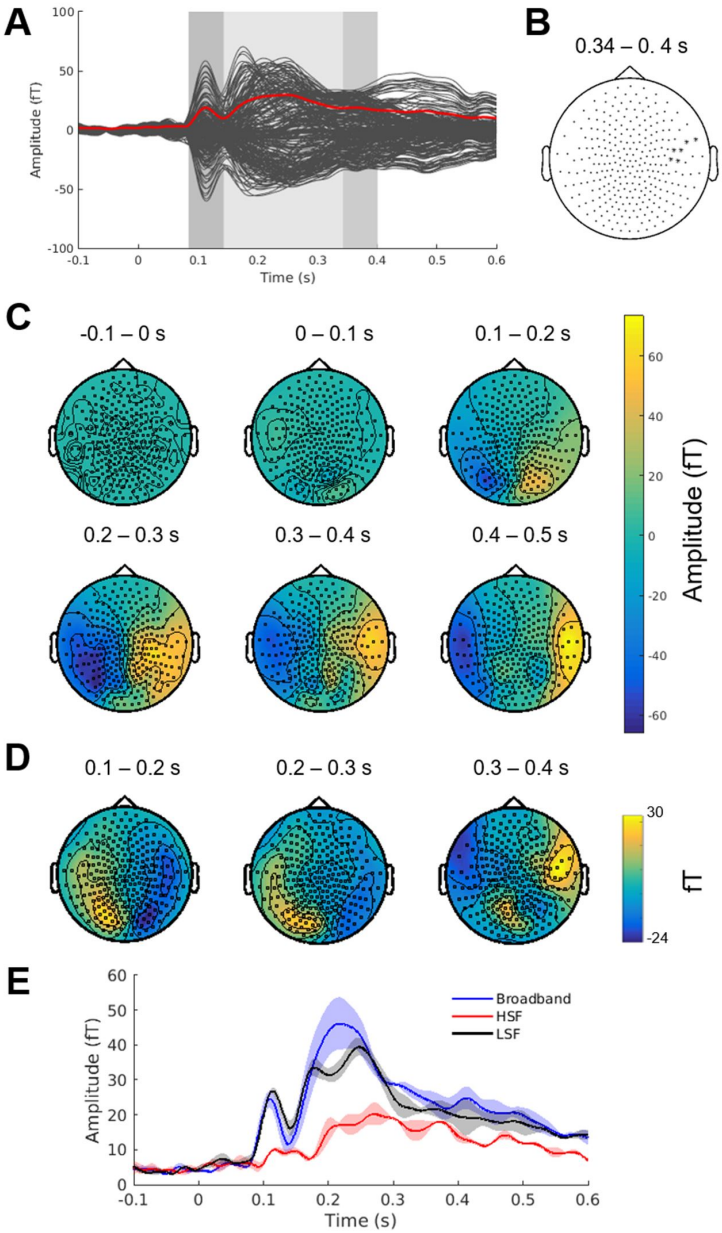

FIGURE 5.7: **A**. Butterfly plot of amplitudes over all trials and all sensors (black) overlaid by the global field power for all trials (red). Local minima in the GFP plot were used to determine windows of interest in the ERF analysis (the shaded gray rectangles represent different time windows). **B**. Sensors exhibiting significant differences in the response to scenes vs scrambled scenes. **C**. Grand average ERF amplitudes in response to unfiltered scenes. **D**. Difference ERF between responses to unfiltered scenes and scrambled stimuli, based on the grand average axial gradient fields. **E**. Grand average global field power for each spatial frequency condition, showing lower amplitude responses to HSF stimuli.

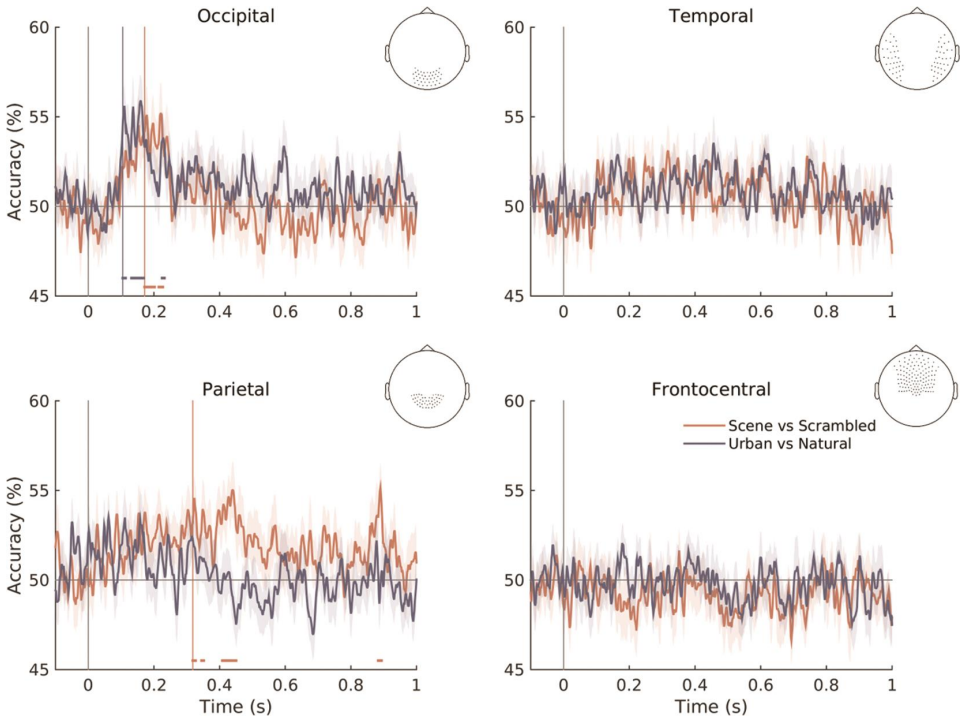

FIGURE 5.8: Unfiltered scene categories: time-resolved decoding accuracy traces ($\pm SEM$) obtained using different sensor sets for both decoding problems. Accuracies were averaged across subjects and smoothed with a five-point moving average for visualization only. Horizontal lines show above-chance decoding performance (P<0.01 corrected).

TABLE 5.1: Scene decoding results in sensor and source space

Scene vs Scrambled	**Occipital**	**Temporal**	**Parietal**	**Frontocentral**	**Source space**
Max accuracy	56.07%	53.93%	56.12%	52.93%	57.41%
95% CI	53.21%, 60.59%	51.14%, 57.4%	53.24%, 59.21%	50.79%, 55.78%	54.4%, 60%
Max sensitivity	58.95%	56.9%	58.38%	54.73%	69.8%
Max specificity	54.96%	53.58%	55.54%	53.34%	57.07%
Decoding onset	172 ms	N/A	318 ms	N/A	215 ms
95% CI	145-215 ms	N/A	83-423 ms	N/A	173-223 ms
Natural vs Urban					
Max accuracy	56.56%	54.55%	54.65%	54.18%	55.91%
95% CI	53.84%, 59.67%	52.13%, 56.94%	51.62%, 57.91%,	51.61%, 56.63%	53.55%, 58.49%
Max sensitivity	57.03%	55.3%	55.07%	52%	55.25%
Max specificity	58.2%	55.72%	56.79%	56.39%	74.06%
Decoding onset	105 ms	N/A	N/A	N/A	105 ms
95% CI	102-146 ms	N/A	N/A	N/A	102-232 ms

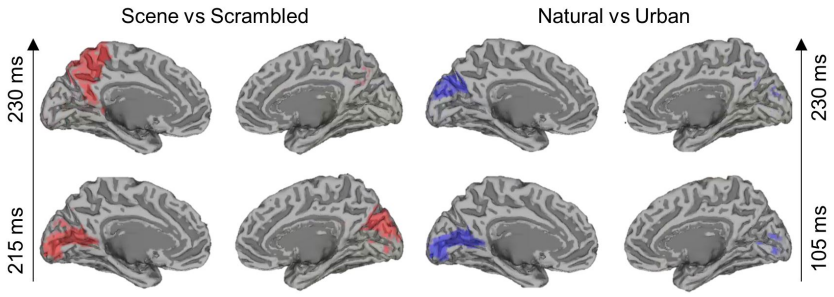

FIGURE 5.9: Unfiltered scene categories: ROIs achieving significant decoding performance across subjects in the searchlight source-space MVPA analysis (P<0.001, cluster-corrected across time).

5.5.2 From low-level to categorical representations

Within-frequency decoding

To assess spatial frequency preferences in the processing of natural scenes, we performed within-spatial frequency and cross-spatial frequency classification using occipital sensor-level MEG responses. Only HSF stimuli achieved above-chance decoding performance in within-spatial frequency classification (Table 5.2). Classification accuracy reached significance at 175 ms post-stimulus onset for the scene vs scrambled decoding problem, and briefly at 183 ms for the urban vs natural scene decoding problem (Figure 5.10), thus following a similar timecourse to the decoding of unfiltered scenes.

Cross-frequency decoding

We performed cross-frequency decoding to evaluate the generalizability of scene responses across spatial frequencies. This allowed us to assess, for example, whether a decoder trained to classify scenes on a set of LSF stimuli could generalize to a set of HSF stimuli and vice versa.

We were unable to detect a truly high-level response (i.e., above-chance generalization across LSF and HSF stimulus sets). Successful cross-decoding was only achieved when classifying between scenes and scrambled stimuli across LSF and broadband stimulus sets (Figure 5.10) starting at ~168 ms after stimulus onset.

Contrast-related asymmetries in SNR pose a potential concern to this analysis (we note lower signal amplitudes in response to high spatial frequency, low

TABLE 5.2: The role of spatial frequency in scene category decoding

	Within-frequency		*Cross-frequency*		
Scene vs Scrambled	**LSF**	**HSF**	**LSF - HSF**	**HSF - Broadband**	**LSF - Broadband**
Max accuracy	54.63%	56.43%	53%	53.42%	55.3%
95% CI	52.24%, 58.73%	54.41%, 58.23%	51.51%, 55.7%	51.96%, 54.75%	53.65%, 57.94%
Max sensitivity	54.47%	55.85%	53.76%	56.66%	56.22%
Max specificity	56.95%	57.31%	54.97%	54.57%	54.97%
Decoding onset	N/A	175 ms	N/A	N/A	168 ms
95% CI	N/A	133-208 ms	N/A	N/A	165-177 ms
Natural vs Urban					
Max accuracy	53.17%	54.97%	52.82%	53.29%	52.91%
95% CI	49.06%, 56.59%	51.41%, 59.41%	51.56%, 54.57%	51.73%, 55.2%	50.89%, 54.62%
Max sensitivity	54.82%	56.38%	54.33%	53.69%	52.58%
Max specificity	55.15%	56.38%	55.75%	56.95%	57.31%
Decoding onset	N/A	183 ms	N/A	N/A	N/A
95% CI	N/A	183-282	N/A	N/A	N/A

contrast stimuli; see Figure 5.7E). However, when decoding scenes from scrambled stimuli within each spatial frequency condition, higher accuracy was achieved on the HSF stimulus set than the higher contrast LSF set (Figure 5.10), suggesting that discriminating information is present at high spatial frequencies despite lower SNR. The lower recognizability of LSF scenes (as shown in the behavioural experiment) may explain the lower accuracies obtained in their classification.

Despite this, cross-decoding results suggest that responses to unfiltered scenes are based on LSF features within 200 ms of stimulus onset. Successful cross-decoding points to a similarly structured multidimensional feature space across conditions, allowing successful generalization of the classifier decision boundary (Grootswagers et al., 2017). In our case, comparable results are achieved in both directions of training and testing, suggesting that despite lower classification rates within the LSF stimulus set, LSF features play an important role in natural scene perception. Although HSF features appear to contain information discriminating scenes from scrambled stimuli, it is more likely that these are associated with low-level perception, as they fail to generalize to broadband scene representations. Together, the MVPA analyses describe natural scene perception as a multi-stage process, with different spatial frequencies playing different roles in the encoding of information in visual cortex.

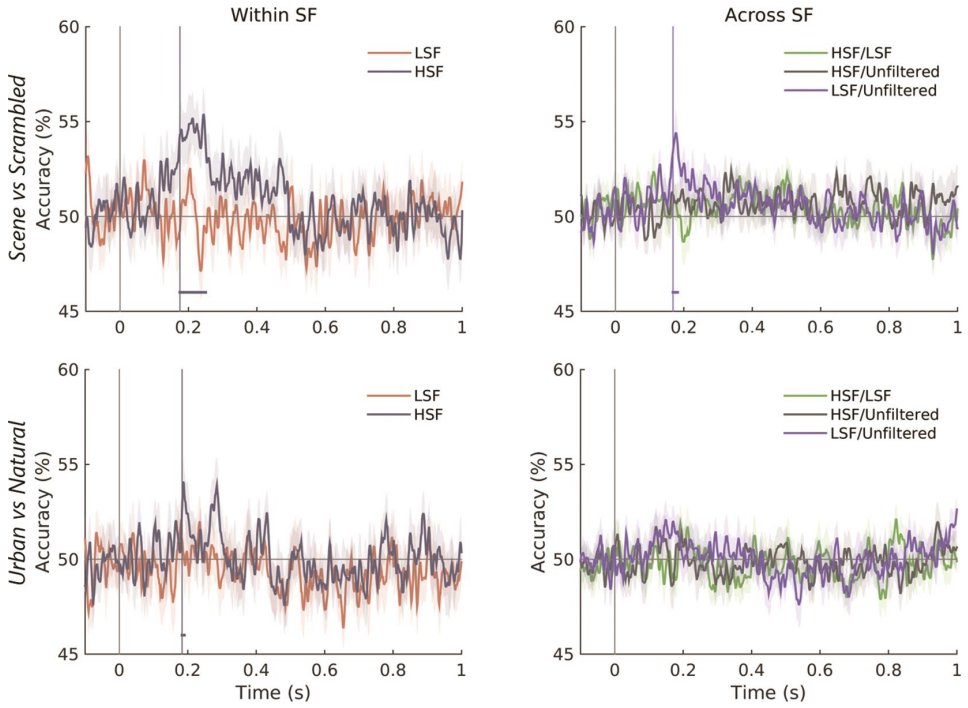

FIGURE 5.10: The role of spatial frequency: Time-resolved decoding accuracies ($\pm SEM$) for both decoding problems using the occipital sensor set. **Left**: decoding within spatial frequency (HSF and LSF); **Right**: cross-decoding across the broadband and LSF stimulus sets. Above-chance decoding time windows are marked with horizontal lines (P<0.01 corrected).

Low-level and categorical representations in visual cortex

We interrogated the structure of neural representations using two RSA analyses. First, we performed RSA to test for partial correlations between MEG responses to scenes and four models guided by low-level properties or high-level category distinctions between stimuli. Neural patterns correlated most often and significantly with the spatial frequency-based model (maximum correlation $\rho = 0.24, P < 0.01$; Figure 5.11), with a few ROIs (shown below) showing significant correlations with the spatial envelope and scene category models (maximum $\rho = 0.18$ and $\rho = 0.14$ respectively, $P < 0.01$). No correlations with the scene identity model reached significance after correction for multiple comparisons ($\rho < 0.16, P > 0.039$).

The spatiotemporal evolution of different scene representations is shown in Figure 5.11B. Before 150 ms, responses in early visual areas such as the lingual gyrus and calcarine cortex significantly correlated with the spatial frequency model, with correlations extending parietally and temporally later (150–250 ms). Interestingly, responses in posterior cingulate, temporal and extrastriate ROIs, where we might

expect selective responses to scenes, correlated with the spatial frequency RDM at relatively late time points. These included areas identified in the MVPA analysis as supporting scene decoding.

Spatial envelope correlations were less represented in this dataset than reported by others (Ramkumar et al., 2016; Rice et al., 2014) and recruited occipito-parietal areas at ~210 ms. Interestingly, these correlations appeared later than those with the scene category model, suggesting overlapping processing of low-, mid- and high-level properties in the visual system (Ramkumar et al., 2016).

While the scene identity model did not predict MEG patterns, the scene category model correlated with responses in the visual cortex at ~180 ms post-stimulus onset. We note that correlations with the spatial frequency and the spatial envelope RDMs were partialled out of this analysis; it is thus likely that these correlations reflect true categorical differences in perception. This stage in processing coincides with the emergence of an occipital LSF scene response in the cross-decoding analysis (Figure 5.10).

After excluding the contribution of the RMS contrast-based RDM from the partial correlation analysis, the spatial frequency sensitivity revealed earlier was diminished. This is in line with previous reports suggesting that spatial frequency processing is dependent on the amplitude spectrum (Andrews et al., 2010; Kauffmann et al., 2015a). RMS contrast also appeared to impact spatial envelope correlations, which arose later in this analysis (Figure 5.12). Interestingly, significant correlations with the category-based model occurred at the same timepoints and in the same ROIs as in the previous analysis, reinforcing the idea that this is a truly high-level response.

While the correlation coefficients are relatively low, with a maximum of 5.7% of the variance explained by the spatial frequency model, the noise estimate suggests that the maximum correlation detectable in our data is low (mean lower and upper bound estimates across time and ROIs of $\rho = 0.038$ and $\rho = 0.25$ respectively). These values are comparable with previous RSA results obtained with similar data (Cichy et al., 2016; Wardle et al., 2016), but higher SNR data (e.g. larger trial numbers) would be desirable to increase sensitivity (Nili et al., 2014)).

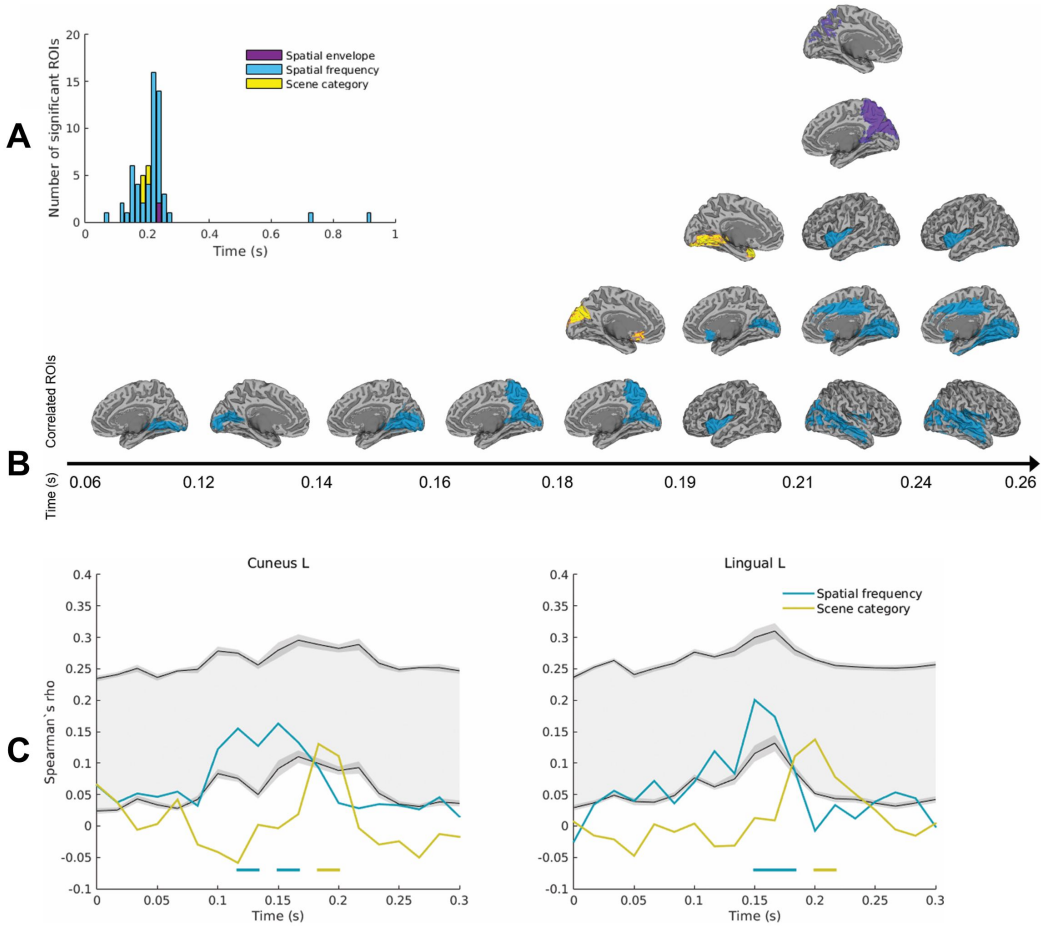

FIGURE 5.11: Unfolding of feature-based model representations. **A**. Number of ROIs significantly correlated with either of the feature-based models over time. **B**. Summary view of the ROIs significantly correlated with either of the feature-based models over time, overlaid on the MNI template brain (P<0.01 corrected). For bilateral ROIs, one hemisphere is shown for clarity. **C**. Example of correlation time-course (in steps of ~16 ms) for the two visual cortex ROIs showing category-related representations. The gray shaded areas represent the noise ceiling, delineated by upper and lower bounds in black. The upper bound was calculated by correlating the average neural RDM across subjects to each individual's neural RDM, while the lower bound was obtained by correlating each individual's RDM to the average of the remaining 18 subjects' RDMs. 95% confidence intervals on the noise ceiling bounds are represented in dark gray. The horizontal lines show significant correlations arising when the correlation coefficient overlaps with the noise estimate, as expected (P<0.01 corrected).

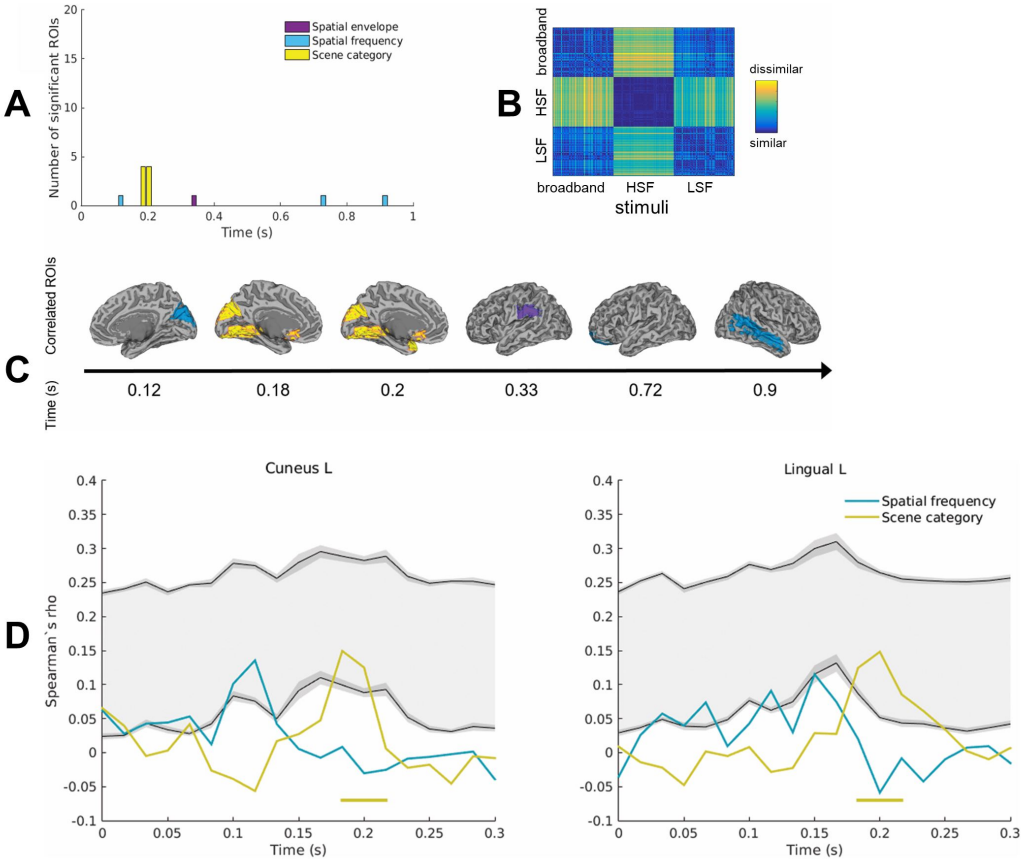

FIGURE 5.12: Unfolding of feature-based model representations after partialling out correlations with RMS contrast. **A**. Number of ROIs significantly correlated with either of the feature-based models over time after partialling out the RMS contrast-based model. **B**. The RMS contrast-based model RDM. **C**. Summary view of the ROIs significantly correlated with either of the feature-based models over time, overlaid on the MNI template brain (P<0.01 corrected). Note that scene category correlations remain virtually unchanged. **D**. Example of correlation time-course for the two ROIs after partialling out RMS contrast.

Overlapping representations of CNN-based models

We performed a second analysis using model RDMs based on layers of a feedforward deep neural network to assess the hierarchy of scene representations in the visual system. Unsurprisingly given the high correlations between layer-specific RDMs (Figure 5.5), only three layers achieved sustained significant partial correlations with the neural patterns: the second convolutional layer (starting at ~80 ms), the first convolutional layer (starting at ~150 ms), and the seventh fully connected layer (~180–200 ms).

In line with the results reported above, these representations were temporally overlapping both in visual cortex and higher-level cortices Figure 5.13). Interestingly, the high-level layer RDM was represented at the same time points as the categorical representations discussed above, but in higher-level areas including the posterior cingulate cortex. This highlights the potential of deep neural networks as a model that can explain representations in scene-selective cortex (as shown by recent fMRI work linking OPA patterns with CNN features: Bonner and Epstein, 2018); however, at ~180–200 ms, both the low-level and high-level CNN layers make significant unique contributions to explaining the variance in these ROIs. Note also that the high-level CNN RDM is correlated to the low-level feature models (Figure 5.5) and is more dependent on stimulus visual properties than the categorical models tested in the previous analysis. Thus, CNN-based representations paint a complementary picture to the feature-based models, while providing additional evidence against a low-to-high hierarchy of scene processing in the visual system.

5.6 Discussion

Using natural and urban scene stimuli filtered at different spatial frequencies, we tracked the spatiotemporal dynamics of scene perception and tested for low-level and high-level representations of scenes using MEG. We report three main findings based on these analyses.

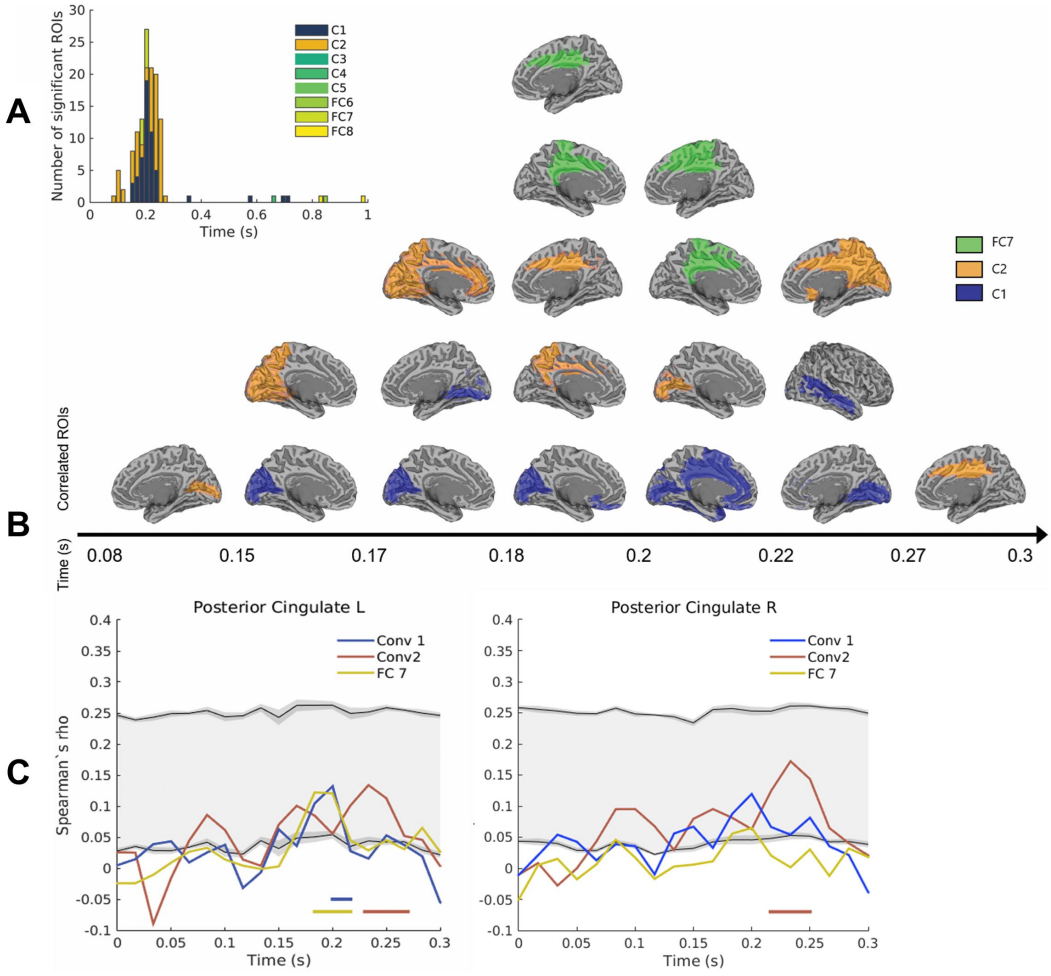

FIGURE 5.13: Unfolding of CNN-based model representations. **A**. Number of ROIs significantly correlated with either of the CNN-based models over time. **B**. Summary view of the ROIs significantly correlated with either of the CNN-based models over time.**C**. Time-course of correlations with CNN-based models in bilateral posterior cingulate cortex.

First, we used MVPA to reveal early (~100 ms) scene processing in the visual cortex. Brain areas along the dorsal and ventral streams encoded information discriminating scenes from scrambled stimuli, while scene category was decodable mainly in visuoparietal cortex.

Second, we used a cross-decoding procedure with independent training and test sets to show the emergence of a response to scenes encoded at low spatial frequencies within 200 ms post-stimulus onset.

Finally, time-resolved RSA results revealed a high-level representation of scene category arising in extrastriate visual cortex at ~180 ms. Both low-level and high-level brain areas contained spatial frequency representations, although these were shown to be dependent on RMS contrast. Furthermore, representations based on layers of a feedforward neural network correlated with patterns in visual and higher-level regions in a temporally overlapping fashion, adding to the evidence of non-hierarchical processing of natural scenes.

5.6.1 Temporal dynamics of scene processing

To date, there has not been extensive electrophysiological research into the temporal dynamics of natural scene processing. Previous studies have isolated responses to scenes by contrasting different types of scenes (Bastin et al., 2013; Cichy et al., 2016; Groen et al., 2013; Groen et al., 2016), or scenes and faces (Rivolta et al., 2012; Sato et al., 1999) or objects (Harel et al., 2016); however, to our knowledge, no previous M/EEG study has used matched control stimuli, which are common in the fMRI literature on natural scenes.

While an early scene-specific event-related field (ERF) component has been reported (M100p: Rivolta et al., 2012), other studies report only late effects (after 200 ms; Groen et al., 2016; Harel et al., 2016; Sato et al., 1999). An MVPA study of natural scenes identified an early low-level response (100 ms) as well as a later signal associated with spatial layout (250 ms; Cichy et al., 2016). Here, we also report evidence of multiple stages in scene processing.

Although no early ERF differences are present in this dataset (possibly due to the matched control stimuli used; Figure 5.7), the MVPA approach revealed single-trial differences starting at ~100 ms for natural vs urban scenes, and at ~170 ms for

scenes vs scrambled stimuli. Classification of natural and urban scenes rose above chance significantly earlier than scene vs scrambled decoding; the occipital origin of this effect suggests a potential contribution of low-level systematic differences between stimulus categories. Successful cross-decoding occurred at similar time points and appeared to reflect a response to scenes based on LSF features, which may be reflected in the simultaneous significant correlations of neural patterns with a scene category model (Figure 5.11). Information about scene category appeared to also be encoded in HSF features at the same time, although this did not generalize across stimulus categories. This response may thus reflect low-level differences encoded at high frequencies and is in line with previous studies showing evidence of responses to HSF images in scene-selective cortex (Berman et al., 2017). Together, these results point to divergent processing of features encoded at different spatial frequencies.

Interestingly, only the extrastriate visual cortex and an area in orbitofrontal cortex showed correlations with categorical scene representations, while the right temporal lobe contained persistent representations of spatial frequency and contrast (Figure 5.11; Figure 5.12). This suggests that visual features may play a part in driving responses in scene-selective areas. This is also supported by overlapping representations of low-level and high-level CNN layer models in areas such as posterior cingulate cortex. On the other hand, categorical responses beyond these areas may be differently represented or may be dependent on behavioural categorization goals.

5.6.2 Mapping scene-selective responses

Extensive fMRI research has mapped responses to natural scenes to the visual cortex, OPA, PPA and RSC (e.g. Nasr et al., 2011; Walther et al., 2009). Here, we used MEG source-space MVPA to detect brain regions responding differently to scenes and scrambled stimuli, or natural and urban scenes respectively. We found differentiating information in visual and parietal cortex when decoding scenes and scrambled stimuli, with more focal patterns discriminating between natural and urban scenes. While the lower sensitivity of MEG to deep sources makes it challenging to detect responses in areas like the PPA, the sources reported here are in

line with previous research reporting occipito-parietal sources of electrophysiological scene-responsive components (Groen et al., 2016; Rivolta et al., 2012).

Furthermore, the RSA mapping of correlations between neural responses and models based on low-level properties or categorical representations showed no classic low-to-high-level dissociation in the visual system. For example, spatial envelope correlations were strongest in occipito-parietal cortex at approximately 230 ms post-stimulus onset, similarly to previously reported correlations with MEG data (Ramkumar et al., 2016), and occurred later than categorical representations. Although not an exhaustive descriptor of scene properties, the spatial envelope model was chosen due to strong evidence that the GIST descriptor accurately represents global scene properties including naturalness, openness, and texture, which match representations in the human visual system (Oliva and Torralba, 2001; Rice et al., 2014; Watson et al., 2017). Significant correlations in parietal areas suggest that scene-specific dorsal stream areas highlighted in the MVPA analysis may rely on image statistics. Finally, neural network representations explained posterior cingulate responses in a temporally and spatially overlapping manner, reinforcing the idea of a complex relationship between visual features and categorical representations.

Spatial frequency and RMS contrast

When contrast was not removed from the RSA analysis, spatial frequency-related representations emerged early (within 100 ms) in the primary visual cortex and extended along the dorsal stream (~160 ms) and later along the ventral stream, as well as parietal and cingulate areas (~200 ms). Despite the limited spatial resolution of MEG and of our ROI-based analysis, we note that correlations were strong in parahippocampal, parietal, cingulate, and inferior occipital areas corresponding to the reported locations of the PPA, RSC and OPA (Figure 5.11). However, when we controlled for RMS contrast, spatial frequency representations only remained strong in visual cortex (~120 ms) and, later, in high-level areas (orbitofrontal and temporal areas; Figure 5.12). This is in line with previous reports showing spatial frequency processing in scene-selective areas (e.g. Nasr et al., 2014; Watson et al.,

2014; Watson et al., 2016, as well as studies suggesting that such effects are dependent on the frequency-specific amplitude spectrum characteristic of natural scenes (Kauffmann et al., 2015a).

Spatial frequency has been previously shown to have a stronger effect on scene recognition than independent contrast manipulation, but the interaction between RMS contrast and spatial frequency elicits the strongest behavioural effects (Kauffmann et al., 2015a). The distribution of contrast across spatial frequency follows a neurobiologically and behaviourally relevant pattern (Andrews et al., 2010; Bex et al., 2009; Guyader et al., 2004), and was maintained in the present study so as to avoid introducing irregularities in the amplitude spectra that would modify natural visual processing strategies. Importantly, contrast did not vary across high-level stimulus categories and only correlated with spatial frequency, ensuring that representations revealed in the MVPA and RSA analyses are contrast-independent.

Categorical representations

In our RSA analysis, category-related representations appeared relatively late in visual cortex, and could be speculatively linked to feedback mechanisms (Peyrin et al., 2010). The proximity of the ROIs to the transverse occipital sulcus suggests the OPA as a potential source of categorical representations.

The emergence of categorical representations at ~180–200 ms post-stimulus onset coincides with previous reports of reaction times in human categorization of natural scenes. Some studies of gist perception report reaction times of at least 250 ms (Rousselet et al., 2005), but studies involving rapid categorization of scenes as natural or man-made interestingly report median reaction times of approximately 200 ms (Crouzet et al., 2012; Joubert et al., 2007). Our data show that at approximately 180 ms the categorical model supersedes the spatial frequency model in visual cortex, while low-level features are simultaneously processed in higher-level areas.

CNN layer representations

Previous research has highlighted the potential of CNNs as powerful models in explaining representations in object- and scene-selective cortex (Groen et al., 2018;

Güçlü and Gerven, 2014; Khaligh-Razavi and Kriegeskorte, 2014; Yamins and DiCarlo, 2016), while an improving understanding of the feature representations employed by CNNs may in turn shed light on the mechanisms underpinning this link (Bonner and Epstein, 2018). In the current study, we extracted layer-specific representations in order to evaluate whether cortical patterns follow the hierarchy of a CNN. We found that high-level CNN representations occurred at the same time as the categorical representations discussed above (and coincided with successful decoding performance in the MVPA analysis). CNN-based models correlated significantly with areas along the dorsal stream, as well as higher-level areas such as the cingulate cortex, with convolutional and fully-connected layers contributing unique information to explaining temporally overlapping cortical patterns.

It is important to reiterate that in both MVPA and RSA analyses, lack of decodable information or significant correlations does not constitute evidence of absence, as information may be otherwise represented in the neural data. However, by comparing multiple models, we provide evidence of the evolution of neural representations in time and space. While the RSA analysis of neural network representations does not match a simple hierachical view of scene processing, it highlights CNN features as good candidate models in explaining scene-selective cortex representations, in line with previous research (Greene and Hansen, 2018; Seeliger et al., 2017; Yamins et al., 2014). On the other hand, the feature-based RSA analysis sees categorical representations arise independently of spatial frequency, RMS contrast, spatial envelope and scene identity, which, unlike the spatial frequency and contrast-based representations, do not involve V1. While early differences in our MVPA analysis may be driven by local low-level differences between scene categories, the RSA analysis points to a later categorical response, simultaneous with the response to low spatial frequencies identified in our cross-decoding analysis.

5.6.3 What's in a category?

A growing body of work suggests that low-level properties play an important part at all stages of processing in the emergence of category-specific representations (Groen et al., 2017). Thus, MVPA analysis results can be difficult to interpret. Even though the stimuli used in our experiment were normalized in terms of contrast

and spatial frequency, a number of properties remain that may differentiate between any two categories, such as the number of edges or the spatial envelope. While it is to be expected that differences in visual properties underpin any differences in high-level representations, assessing the role of low-level properties can help elucidate the source of pattern differences found in our study. Thus, the cross-decoding and RSA analyses provide additional evidence of a categorical stage in natural scene perception and help differentiate this from the earlier, visually driven response revealed by MVPA.

The passive viewing paradigm employed here approached natural viewing conditions and ensured that category effects were not driven by task-related processing, while still controlling for low-level confounds. In the absence of a categorization task, we failed to detect a truly high-level response in our cross-decoding analysis (i.e., generalization across low and high frequency stimuli; Figure 5.10). However, the scene-specific response revealed in the decoding analysis generalized across unfiltered and LSF stimuli within 200 ms, suggesting that low frequency cues encode scene-specific information at later stages of scene processing. Future studies could apply a cross-decoding procedure to data collected using a categorization task in order to investigate the presence of a frequency-invariant response.

Furthermore, we note that failure to achieve above-chance decoding performance in LSF decoding or cross-decoding does not preclude the existence of differential responses that are otherwise represented in the brain, or that the current study design did not detect. However, the current results are informative in comparing conditions and linking the decodability of stimulus categories to spatial frequency information, thus pointing to preferences in spatial frequency processing that may underpin the rapid perception of natural scenes.

Although the repetition of a limited set of stimuli across different spatial frequencies has advantages in terms of controlling for low-level properties, this also poses the concern of stimuli being recognizable between spatial frequency conditions, thus potentially affecting the category differences observed here. However, the fact that we were unable to cross-decode LSF and HSF scenes suggests that such a recognition response could not have significantly contributed to decoding results. Furthermore, such recognition would be expected to affect all conditions equally

(given the stimulus randomization procedure), and would therefore not explain the spatial frequency-specific effects reported here. Finally, we included a scene identity model RDM in our feature-based RSA analysis to assess the recognition of individual scenes across spatial frequency conditions and found no significant correlations with the neural patterns. However, future studies could alleviate this concern by including a larger number of stimuli.

Scene perception is understood as involving a coarse-to-fine processing sequence using both low spatial frequency cues (rapidly processed and allowing for parsing of global structure) and high frequency information (which is relayed more slowly to high-level areas; Kauffmann et al., 2014). The results described in this chapter link the processing of low frequency cues to the formation of categorical representations, supporting previous reports of coarse visual analysis as rapid and crucial to gist perception (Kauffmann et al., 2017; Peyrin et al., 2010; Schyns and Oliva, 1994). On the other hand, HSF representations of scenes do not generalize to unfiltered stimuli, suggesting that they may encode low-level differences rather than a categorical response. However, the presence of such a response may reflect HSF representations previously found in visual and scene-selective areas (Berman et al., 2017; Walther et al., 2011).

Behavioural results obtained through a separate experiment revealed that scenes filtered at low spatial frequencies are more difficult to distinguish from scrambled stimuli than unfiltered or highpass-filtered scenes. This difference was reflected in the lower decodability of LSF scenes from scrambled stimuli. Low-frequency scenes thus appear to be more similar to their scrambled counterparts; interestingly, the similarity in contrast between low-frequency and unfiltered scenes does not provide a categorization or decoding advantage.

However, the difference between the categorization task in the behavioural experiment, with its speed/accuracy tradeoff, and the passive viewing paradigm used in the MEG, means that behavioural results need to be interpreted cautiously. The high behavioural performance across participants (over 90%) suggests that despite these differences, stimuli were generally recognizable across categories.

Challenging ideas of a low-to-high-level hierarchy in the visual system, recent

studies have emphasized the role of low-level properties in scene-selective perception, while at the same time suggesting that categorical distinctions play an important role in behavioural decision-making (Rice et al., 2014; Watson et al., 2016). Such distinctions may emerge from image features and are not "explained away" by low-level properties (Groen et al., 2017; Watson et al., 2017). This chapter takes a step further in explaining how high-level representations arise from the processing of visual features. The RSA and cross-decoding results suggest that spatial frequency is relevant in scene perception, with low-frequency features carrying the information identifying natural scenes as such. Within 200 ms, human visual cortex patterns switch from a low-level representation of stimuli to a categorical representation independent of spatial frequency, contrast and spatial envelope. Furthermore, a convolutional neural network explains representations in visual and cingulate cortex, with high-level layers being represented within 200 ms. These representations arise in the absence of a task, highlighting the remarkable efficiency with which features are extracted from our environment.

The account of visual perception emerging from these results is thus complementary to the conclusions drawn from experiments involving emotional faces in previous chapters. Categorization of highly relevant stimuli is reflected by neural patterns even when this is not the object of a task (Chapter 2), while behavioural goals can impact the spatiotemporal dynamics of visual processing (Chapter 3). Rather than linearly transforming features into concepts, the visual system enables human cognition by performing relevance-based selection of cues from the earliest stages of perception, based on both current behavioural goals, and the evolutionarily adaptive salience of faces and places.

Chapter 6

General discussion

This thesis investigated expression and scene perception with MEG and multivariate analysis tools. Information mapping techniques revealed the spatiotemporal dynamics of perceptual processing, starting with category-related biases at the earliest stages of vision. Three of the main ideas about high-level vision suggested by these findings are highlighted below and are discussed in more detail in the following sections:

1. Although the chapters on face perception highlight the rapid processing of faces and expressions, there is no evidence of an expression-related MEG response outside awareness and attention. However, expression explains some of the variance in behavioural responses made outside awareness.

2. Categorical divisions in the ventral visual system are likely to be explained by feature-based representations linking stimulus properties and behavioural/ conceptual constructs.

3. The temporal and representational dynamics underpinning categorization adapt to task demands and viewing conditions, maximizing efficiency in line with behavioural goals.

6.1 Summary of the findings

The first three chapters of this thesis offer complementary perspectives on expression processing. In Chapter 2, using a passive-viewing paradigm, an early differential response to angry faces compared to happy and neutral faces, originating in

the visual cortex, was found and is likely to represent a modulation of the feedforward response due to emotional salience. In contrast, when using an explicit expression discrimination task, Chapter 3 found early expression-specific processing regardless of valence. Neural patterns in the occipitotemporal cortex encoded face configuration and correlated with behavioural responses; furthermore, the temporal dynamics of feature processing varied with face presentation duration. Finally, a MEG response to faces presented outside of awareness was detected, but no such response was found to subliminal expressions, despite the presence of a behavioural effect. Similarly, Chapter 4 found no responses to expression outside of attention, when emotional faces were presented as distractors during an orientation discrimination task. This did not depend on task difficulty, suggesting that emotional faces were not differentially processed when covert attention was directed to the opposite hemifield.

In the final chapter, responses to natural scenes were evaluated during passive viewing. Low-level visual features and scene categories were processed in a temporally overlapping fashion, with a categorical response to scenes emerging at ~180 ms in the extrastriate visual cortex.

6.2 Categorical responses in passive viewing

Two chapters in the thesis investigated category-related responses in passive viewing. Chapter 2 investigated MEG responses to facial expressions and scrambled stimuli, while Chapter 5 assessed the role of spatial frequency and scene category in natural scene perception.

Despite the lack of a categorization task, both chapters show that stimulus category biases visual processing. In Chapter 2, angry expressions elicit early differential patterns (within 100 ms) in visual cortex, suggesting a role for bottom-up emotional salience in passive viewing. In Chapter 5, categorical effects in scene perception emerge at ~180 ms, in line with other reports on scene categorization (De Cesarei et al., 2018; Mohsenzadeh et al., 2018a; Ramkumar et al., 2016). In the absence of a task, this highlights a system optimized for extracting category-related information. However, temporally overlapping representations of low-level and

high-level information suggest a more complex interplay between features and categories.

Although in Chapter 2 no analyses of representational content could be conducted, the controlled stimulus set and inverse pattern of effects observed in control analyses validates the results as likely to reflect early bottom-up processing of salient cues. This account is supported by the source-space analysis, linking early effects to visual cortex activity and later stages to temporal and frontal regions. In future studies, it would be interesting to test whether the features supporting expression processing under passive viewing are different from the features extracted during an expression discrimination task (Chapter 3).

6.3 Expression processing outside awareness and attention

In Chapter 3 and Chapter 4, participants viewed the same controlled set of emotional face stimuli under different task conditions: a rapid expression discrimination task with some targets shown outside subjective awareness, and a covert spatial attention task in which emotional faces acted as distractors. While all emotional expressions were differentially processed as early as ~100 ms when they were the object of a task, no evidence of differential neural responses to expression processing was found outside of awareness and attention.

In line with results from Chapter 2, Chapter 3 showed early expression processing (~100 ms). This effect emerged only 10-20 ms later than face processing, suggesting that the extraction of features involved in expression detection occurs at the early stages of vision. Furthermore, although the two experiments cannot be directly compared, performing an expression-related task ensured early differential processing of all three expressions, as opposed to the passive viewing context of Chapter 2.

In Chapter 3, expression perception was associated with the extraction of face features and configuration in the ventral strean, when successfully represented (within awareness). Furthermore, a face detection response localized to occipital and ventral areas was found outside awareness. This supports a multi-stage account of face perception, with separable face detection and analysis mechanisms,

despite the near-simultaneous decoding latencies of faces and expressions found within awareness. Face detection may rely on coarse, rapidly extracted contrast structure that is specific to faces without containing any detail (Sinha, 2002).

Although no such response was found to expressions presented outside awareness, expression explained a non-negligible portion of the variance in behaviour. This suggests that the processing of expression outside awareness is supported by qualitatively different neural mechanisms. For example, it is important to note that these data may have been suboptimal for the detection of a subcortical response, especially if a response to expression outside awareness were to recruit the amygdala via a subcortical route. A possible effect might have been too weak to be detected in a whole-brain analysis, especially in deep structures, and the spatial analysis focused on the ventral visual stream specifically, which only represented face information within awareness.

Finally, in Chapter 4, no evidence of expression processing outside attention was found with an identical set of face stimuli, presented as distractors during an orientation discrimination task. Although in this chapter the hypothesis of an intact face detection response could not be tested, the results are in line with reports suggesting that expression perception requires attention, or at least sufficient attentional resources (Chen et al., 2016; Devue and Grimshaw, 2017; Puls and Rothermund, 2018).

Together, these results highlight the importance of both bottom-up and top-down influences in expression processing, and suggest that the visual system is highly adaptable to different viewing conditions and behavioural goals. The pattern of results across the three chapters on face perception highlights above all the highly dynamic nature of the visual system, and its amenability to both endogenous and exogenous factors.

6.4 Axes in representational space

Decoding analyses typically assess the presence of discriminating information in neural patterns, while cross-decoding and temporal generalization approaches investigate their invariance to irrelevant features and their temporal structure. However, a different approach is needed in order to evaluate the type of information encoded in these patterns, especially when different stimulus properties may correlate with their category (Chapter 1). In Chapter 3 and Chapter 5, conflicting hypotheses about the featural or categorical nature of MEG patterns were tested using RSA.

In any model-testing approach, one of the main challenges is restricting the possible model space, while at the same time exploring possible alternatives and confounds. To balance hypothesis testing and data-driven exploration (Kriegeskorte and Kievit, 2013), it is important to test a range of models based on prior information and qualitatively different theories. Comparing models and linking them to behaviour can help refine hypotheses and pinpoint properties essential in representation across the visual stream. In the RSA analyses described here, different featural and categorical models based on previous findings are compared, together with properties orthogonal to the categories of interest.

In Chapter 3, a set of 9 models was tested using a spatiotemporal searchlight approach in occipitotemporal cortex, including a first-order and second-order face configuration model (Diamond and Carey, 1986), spatial envelope, and expression and identity models. The time-resolved approach revealed a sequence of stages in face configuration processing with the potential to explain previous conflicting accounts; it also highlighted the dynamic adaptability of such processes to changes in visual input, and their link to behavioural responses. These results support the idea of featural coding of face category in the ventral stream (Bracci et al., 2017a) undergoing continuous optimization in response to behavioural goals, context and stimulus properties. In this chapter, the behavioural model was particularly informative in assessing the relevance of both neural patterns and stimulus features. Establishing a link between behaviour and neural representations can suggest that

the latter are actively used in processing (Carlson et al., 2017), and is thus an important step when studying task-related processes.

A similar picture emerges from the results of Chapter 5, which contradict strictly hierarchical models for visual processing. Ongoing low-level feature processing appears to underscore categorical representations in both visual and temporal areas. Furthermore, directly testing hierarchical feature representations using layer activations from a DNN reveals additional evidence: although high-level layers are represented mainly in posterior cingulate patterns, and later than low-level layers (~200 ms), low-level layers elicit temporally overlapping correlations with patterns in both visual and higher-level brain areas.

Although models based on stimulus properties and those based on a DNN represent the stimuli in different ways, the best performing among both sets explain similar amounts of variance and reach the noise ceiling. This highlights, once again, the importance of building plausible models, informed by theory, biology, behaviour, or previous findings, and reminds us of the correlational nature of evidence in neuroimaging analyses. In this case, both approaches have their advantages, and their convergence might ultimately prove to hold the most explanatory power. Models based on stimulus properties are simple and understandable, have direct counterparts in behaviour and psychological concepts, and can point us in the right direction when evaluating opposing hypotheses about the level of abstraction employed by neural coding in the visual system. On the other hand, DNNs could inform our hypotheses with new and efficient representations of the data along category axes. Although at the moment this possibility is limited by the complexity and biological implausibility of DNNs in their current form, architectures inspired from neural circuits and optimization of DNNs using naturalistic stimuli and tasks could alleviate this problem. Furthermore, a better understanding of the operations performed by DNNs and which of these could be implemented by neural codes would help the two approaches converge.

Uncovering the axes of representational spaces in the brain is a difficult problem, in part due to the large number of possible solutions; however, this can be allievated by approaches like RSA, based on abstracting to a level of representation where hypotheses become testable, and by computational models bringing

together advances in computer vision and biological constraints.

6.5 Towards dynamic representations

Selection mechanisms are essential in visual processing, given the complexity of visual input: an exhaustive analysis of the environment would be inefficient. Visual perception can thus be seen as a type of perceptual decision-making (Seger and Peterson, 2013), with both bottom-up and top-down modulations contributing to visual analysis. In line with this view, one of the main ideas emerging consistently from all chapters is that of a dynamic, efficient, and adaptable visual system, which both detects evolutionarily relevant cues in the environment, and optimizes its responses in accordance with behavioural goals. In Chapter 2 and Chapter 5, we see evidence of the former in the early expression-related biases and the emergence of categorical responses in the absence of a task. Conversely, in Chapter 3 and Chapter 4, we see top-down effects impact expression processing and temporal dynamics change in order to accomplish the same goal under different conditions. In keeping with recent discussions of how top-down factors shape perception at all stages (Gilbert and Li, 2013), these results suggest that the search for a neural representational code may have been made more difficult by its dynamic characteristics.

The classic view of visual recognition as a hierarchical process implemented in the ventral stream (DiCarlo et al., 2012) has recently been challenged by studies demonstrating how task demands change stimulus representations (Bracci et al., 2017b; Erez and Duncan, 2015; Hebart et al., 2018; Nastase et al., 2017; Vaziri-Pashkam and Xu, 2017). Although ventral stream representations are thought to be less affected by task demands (Bracci and Beeck, 2016; Vaziri-Pashkam and Xu, 2017), few studies have investigated the temporal dynamics of these effects. In Chapter 3, changes in available visual information are reflected in the temporal dynamics of ventral feature representations more prominently than in their spatial extent; this highlights the importance of studying such processes with high temporal resolution. Furthermore, evidence of changing information content within the same cortical areas (Vida et al., 2017, Chapter 3, Chapter 5) points to the dynamic nature of neural representations, likely to be supported by activation patterns that

rapidly change in response to feedforward and feedback information. As the results in this thesis seem to suggest, combining task manipulations with controlled visual stimuli and time-resolved multivariate techniques is likely to reveal a complex picture of the adaptable neural code supporting high-level vision.

6.6 Conclusions and future directions

6.6.1 Multivariate analyses

This thesis demonstrated the strength of multivariate analysis approaches in tracking the dynamics of perceptual processing, and showed how source-space decoding and model testing approaches can tease apart different ideas about neural representations. From the outset, multivariate methods were used not for prediction, but for interpretation (Hebart and Baker, 2017), and analysis choices were made to maximize interpretability. Combining sensor-level with source-space approaches offered complementary information about the spatiotemporal correlates of visual perception; cross-exemplar decoding captured categorical responses; cross-decoding across time and conditions tested the invariance of these responses; and finally, randomization testing assessed the presence of information against an empirical null distribution. Using controlled experimental designs and stimulus sets was also essential in maximizing the interpretability of these results.

The experiments in this thesis highlight the need to account for low-level properties and other irrelevant features that may covary with the category of interest, especially when using sensitive multivariate methods. Methods like RSA permit the explicit modelling of features that might be contributing to the signal. However, the selection of appropriate control models can be challenging, as both feature-based and computational models have been used as proxies for low-level features. In Chapter 3 and Chapter 5, a feature-based approach evaluates the unique variance explained by different stimulus properties. An advantage of this approach is that the timecourse and spatial extent of low-level features such as contrast is relatively well-predicted by existing knowledge, which makes them suitable controls; this makes results easier to evaluate than when using less explicit computational models. Furthermore, throughout the thesis, a stimulus normalization approach is used

in addition to feature modelling. For example, in Chapter 3, the alignment, cropping, contrast matching and Fourier amplitude normalization of the faces leave only variance in local features uncontrolled. Although the best approach depends on the aims of each study and the trade-off between naturalistic and controlled stimuli, such considerations are particularly important when employing multivariate methods.

Furthermore, in most analyses, no assumptions were made about the timing or localization of an effect, and any ROI selection was performed using data-driven approaches (e.g. by localizing responses to faces). Together, spatiotemporally-resolved decoding and RSA analyses comprehensively described both the dynamics of visual perception and their representational content. Although not a standard approach, performing space-resolved RSA of MEG data takes advantage of the localization capabilities of MEG and can be a good alternative to cross-modality RSA. In this thesis, data-driven analyses successfully resolved perceptual processing in space and time and linked it to computational models and behaviour. As large-scale, collaborative, cross-modal datasets become more and more common, data-driven tools can help make sense of rich information and reveal shared patterns (Baillet, 2017; Smith and Nichols, 2018).

Given the rapid progress of analysis methods and the high dimensionality of MEG datasets, there are many possible choices in the multivariate analysis of MEG data. It is only recently that methodological studies have started to uncover the strengths and weaknesses of some commonly used metrics and approaches (Guggenmos et al., 2018; Sato et al., 2018). Future work will use simulated data and the dataset described in Chapter 5 to quantitatively assess the impact of using different analysis pipelines, particularly in source-space decoding. As methods are being improved, the interpretability and versatility of multivariate analyses will also increase; for example, cross-exemplar decoding, cross-validation, and noise normalization procedures are becoming widely adopted, improving the reliability of decoding results.

6.6.2 Clinical relevance

One of the most appealing characteristics of multivariate methods is their potential to offer a sensitive, automated and individual-specific marker of neural processing. As such, decoding methods have been widely implemented in brain-computer interfacing (Horschig et al., 2015), neurofeedback (Okazaki et al., 2015), and in clinical research (Lu et al., 2013). Although this thesis focused on decoding for understanding brain function in healthy populations, there is a potential for expanding this work for clinical research. For example, Chapter 3 shows rapid decoding of expressions presented as briefly as 150 ms; such a rapid presentation paradigm can be implemented in patient populations to investigate the neural markers associated with differences in expression discrimination ability (e.g. Clark and Mcintosh, 2008; Kohler et al., 2011; Riwkes et al., 2015). Future work will focus on establishing the potential of within-subject information metrics such as decoding accuracy in quantifying individual differences, and look at assessing this in patient populations.

6.6.3 Future directions

Methodological advances in machine learning, technological advances like on-scalp MEG, and the increase in large-scale collaborations and data sharing signal an exciting time in cognitive neuroscience. Together, these factors can increase the sensitivity and spatiotemporal resolution of non-invasive measures of neural activity. At the same time, hypothesis-generating computational models may link neural, psychological, and behavioural levels of analysis within one framework, in which the building blocks of perceptual processing are representations emerging within dynamic neural circuits.

As we move from a hierarchical, object recognition framework of high-level vision to a dynamic model of feature-based representations, finding the right model will require a combination of experimental designs testing the boundaries of this adaptability, and computational tools optimized for specific goals. The successful implementation of object recognition in artificial systems suggests that an understanding of vision is not out of reach for modern computational models; however,

the dynamic aspect of brain computations, which is increasingly highlighted as important for an efficient and sparse neural coding strategy, is missing in most machine learning algorithms (VanRullen, 2017). Starting with naturalistic task-based optimization, or adopting theoretical neuroscience frameworks such as predictive coding (Hassabis et al., 2017; Rawlinson and Kowadlo, 2017), might be some ways of increasing the biological plausibility of machine learning algorithms. In turn, this could open the way to tackling problems beyond object recognition: cognition, emotion and social perception could be integrated within an information processing framework. Although we are far from an understanding of vision in Marr's terms, machine learning offers a new testing ground for potential strategies employed by the brain to achieve successful, task-relevant visual categorization.

Bibliography

Adams, W. J. et al. (2010). "High-Level Face Adaptation Without Awareness". In: *Psychological Science* 21.2, pp. 205–210. DOI: 10.1177/0956797609359508.

Aguado, L. et al. (2012). "Modulation of early perceptual processing by emotional expression and acquired valence of faces: An ERP study". In: *Journal of Psychophysiology* 26.1, pp. 29–41. DOI: 10.1027/0269-8803/a000065.

Almeida, J. et al. (2013). "Affect of the unconscious : Visually suppressed angry faces modulate our decisions". In: *Cognitive, Affective, & Behavioral Neuroscience* 13, pp. 94–101. DOI: 10.3758/s13415-012-0133-7.

Andrews, T. J. et al. (2010). "Selectivity for low-level features of objects in the human ventral stream". In: *NeuroImage* 49.1, pp. 703–711. DOI: 10.1016/j.neuroimage.2009.08.046.

Andrews, T. J. et al. (2015). "Low-level properties of natural images predict topographic patterns of neural response in the ventral visual pathway visual pathway". In: *Journal of Vision* 15.7, pp. 1–12. DOI: 10.1167/15.7.3.doi.

Axelrod, V. (2010). "The Fusiform Face Area: In Quest of Holistic Face Processing". In: *Journal of Neuroscience* 30.26, pp. 8699–8701. DOI: 10.1523/JNEUROSCI.1921-10.2010.

Axelrod, V., M. Bar, and G. Rees (2015). "Exploring the unconscious using faces". In: *Trends in Cognitive Sciences* 19.1, pp. 35–45. DOI: 10.1016/j.tics.2014.11.003.

Bahramisharif, A. et al. (2012). "The dynamic beamformer". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 7263 LNAI, pp. 148–155. DOI: 10.1007/978-3-642-34713-9{_}19.

Baillet, S. (2017). "Magnetoencephalography for brain electrophysiology and imaging". In: *Nature Neuroscience* 20.3, pp. 327–339. DOI: 10.1038/nn.4504.

Baillet, S., J. C. Mosher, and R. M. Leahy (2001). "Electromagnetic brain mapping". In: *IEEE Signal Processing Magazine* 18.6, pp. 14–30. DOI: 10.1109/79.962275.

Balconi, M. and U. Pozzoli (2003). "Face-selective processing and the effect of pleasant and unpleasant emotional expressions on ERP correlates". In: *International Journal of Psychophysiology* 49.1, pp. 67–74. DOI: 10.1016/S0167-8760(03)00081-3.

Baldassi, C. et al. (2013). "Shape Similarity, Better than Semantic Membership, Accounts for the Structure of Visual Object Representations in a Population of Monkey Inferotemporal Neurons". In: *PLoS Computational Biology* 9.8. DOI: 10.1371/journal.pcbi.1003167.

Baltrusaitis, T., P. Robinson, and L.-P. Morency (2016). "OpenFace: an open source facial behaviour analysis toolkit". In: *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1–10.

Bankson, B. B. et al. (2018). "The temporal evolution of conceptual object representations revealed through models of behavior, semantics and deep neural networks". In: *NeuroImage* 178, pp. 172–182. DOI: 10.1016/J.NEUROIMAGE.2018.05.037.

Bar, M et al. (2006). "Top-down facilitation of visual recognition". In: *Proceedings of the National Academy of Sciences* 103.2, pp. 449–454. DOI: 10.1073/pnas.0507062103.

Barnes, G. R. et al. (2004). "Realistic spatial sampling for MEG beamformer images". In: *Human Brain Mapping* 23.2, pp. 120–127. DOI: 10.1002/hbm.20047.

Bastiaansen, M. C. and T. R. Knösche (2000). "Tangential derivative mapping of axial MEG applied to event-related desynchronization research." In: *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology* 111.7, pp. 1300–5.

Bastin, J. et al. (2013). "Timing of posterior parahippocampal gyrus activity reveals multiple scene processing stages". In: *Human Brain Mapping* 34.6, pp. 1357–1370. DOI: 10.1002/hbm.21515.

Beck, J. (2018). "Marking the Perception–Cognition Boundary: The Criterion of Stimulus-Dependence". In: *Australasian Journal of Philosophy* 96.2, pp. 319–334. DOI: 10.1080/00048402.2017.1329329.

Beeck, H. P. Op de, K. Torfs, and J. Wagemans (2008). "Perceived Shape Similarity among Unfamiliar Objects and the Organization of the Human Object Vision Pathway". In: *Journal of Neuroscience* 28.40, pp. 10111–10123. DOI: 10.1523/JNEUROSCI.2511-08.2008.

Behrmann, M. et al. (2014). "Holistic Face Perception". In: *Oxford Handbook of Perceptual Organization*. Ed. by J. Wagemans. Oxford: Oxford University Press, pp. 1–21. DOI: 10.1093/oxfordhb/9780199686858.013.010.

Bell, A. H. et al. (2009). "Object representations in the temporal cortex of monkeys and humans as revealed by functional magnetic resonance imaging." In: *Journal of neurophysiology* 101.2, pp. 688–700. DOI: 10.1152/jn.90657.2008.

Belopolsky, A. V., C. Devue, and J. Theeuwes (2011). "Angry faces hold the eyes". In: *Visual Cognition* 19.1, pp. 27–36. DOI: 10.1080/13506285.2010.536186.

Ben-Hur, A. and J. Weston (2010). "A User's Guide to Support Vector Machines". In: *Data Mining Techniques for the Life Sciences. Methods in Molecular Biology (Methods and Protocols), vol 609*. Humana Press, pp. 223–239. DOI: 10.1007/978-1-60327-241-4{_}13.

Berman, D., J. D. Golomb, and D. B. Walther (2017). "Scene content is predominantly conveyed by high spatial frequencies in scene-selective visual cortex". In: *PLoS ONE* 12.12, pp. 1–16. DOI: 10.1371/journal.pone.0189828.

Bernstein, M. and G. Yovel (2015). "Two neural pathways of face processing: A critical evaluation of current models". In: *Neuroscience and Biobehavioral Reviews* 55, pp. 536–546. DOI: 10.1016/j.neubiorev.2015.06.010.

Bertini, C., R. Cecere, and E. Làdavas (2017). "Unseen fearful faces facilitate visual discrimination in the intact field". In: *Neuropsychologia* June, pp. 1–71. DOI: 10.1016/j.neuropsychologia.2017.07.029.

Bex, P. J., S. G. Solomon, and S. C. Dakin (2009). "Contrast sensitivity in natural scenes depends on edge as well as spatial frequency structure." In: *Journal of vision* 9.10, pp. 1–19. DOI: 10.1167/9.10.1.

Bishop, S. et al. (2004). "Prefrontal cortical function and anxiety: controlling attention to threat-related stimuli". In: *Nature Neuroscience* 7.2, pp. 184–188. DOI: 10.1038/nn1173.

Boehler, C. N. et al. (2008). "Rapid recurrent processing gates awareness in primary visual cortex". In: *Proceedings of the National Academy of Sciences* 105.25, pp. 8742–8747.

Bonner, M. F. and R. A. Epstein (2017). "Coding of navigational affordances in the human visual system". In: *Proceedings of the National Academy of Sciences* 114.18, pp. 4793–4798. DOI: 10.1073/pnas.1618228114.

Bonner, M. F. and R. A. Epstein (2018). "Computational mechanisms underlying cortical responses to the affordance properties of visual scenes". In: *PLoS Computational Biology* 022350, pp. 1–31.

Boser, B. E., I. M. Guyon, and V. N. Vapnik (1992). "A training algorithm for optimal margin classifiers". In: *Proceedings of the fifth annual workshop on Computational learning theory - COLT '92*. New York, New York, USA: ACM Press, pp. 144–152. DOI: 10.1145/130385.130401.

Boto, E. et al. (2017). "A new generation of magnetoencephalography: Room temperature measurements using optically-pumped magnetometers". In: *NeuroImage* 149.January, pp. 404–414. DOI: 10.1016/j.neuroimage.2017.01.034.

Boto, E. et al. (2018). "Moving magnetoencephalography towards real-world applications with a wearable system". In: *Nature* 555.7698, pp. 657–661. DOI: 10.1038/nature26147.

Bracci, S. and H. Op de Beeck (2016). "Dissociations and Associations between Shape and Category Representations in the Two Visual Pathways." In: *The Journal of Neuroscience* 36.2, pp. 432–44. DOI: 10.1523/JNEUROSCI.2314-15.2016.

Bracci, S., J. B. Ritchie, and H. O. de Beeck (2017a). "On the partnership between neural representations of object categories and visual features in the ventral visual pathway". In: *Neuropsychologia* 105, pp. 153–164. DOI: 10.1016/j.neuropsychologia.2017.06.010.

Bracci, S., N. Daniels, and H. Op de Beeck (2017b). "Task Context Overrules Object- and Category-Related Representational Content in the Human Parietal Cortex". In: *Cerebral Cortex* 27.1, pp. 310–321. DOI: 10.1093/cercor/bhw419.

Bradley, B. P., K. Mogg, and N. H. Millar (2000). "Covert and overt orienting of attention to emotional faces in anxiety". In: *Cognition and Emotion* 14.6, pp. 789–808. DOI: 10.1080/026999300501566636.

Brainard, D. H. (1997). "The Psychophysics Toolbox". In: *Spatial Vision* 10, pp. 433–436. DOI: http://dx.doi.org/10.1163/156856897X00357.

Brookes, M. J. et al. (2016). "A multi-layer network approach to MEG connectivity analysis". In: *NeuroImage* 132, pp. 425–438. DOI: 10.1016/j.neuroimage.2016.02.045.

Bruce, V. and A. Young (1986). "Understanding face recognition". In: *British Journal of Psychology* 77, pp. 305–327.

Bullier, J (2001). "Integrated model of visual processing". In: *Brain Research Reviews* 36, pp. 96–107. DOI: 10.1016/S0165-0173(01)00085-6.

Caldara, R. et al. (2006). "The fusiform face area is tuned for curvilinear patterns with more high-contrasted elements in the upper part". In: *NeuroImage* 31.1, pp. 313–319. DOI: 10.1016/j.neuroimage.2005.12.011.

Calder, A. J. and A. W. Young (2005). "Understanding the recognition of facial identity and facial expression." In: *Nature Reviews Neuroscience* 6.8, pp. 641–51. DOI: 10.1038/nrn1724.

Calder, A. J. et al. (2000). "Configural information in facial expression perception." In: *Journal of Experimental Psychology: Human Perception and Performance* 26.2, pp. 527–551. DOI: 10.1037/0096-1523.26.2.527.

Calvo, M. G., P. Avero, and D. Lundqvist (2006). "Facilitated detection of angry faces: Initial orienting and processing efficiency". In: *Cognition and Emotion* 20.6, pp. 785–811. DOI: 10.1080/02699930500465224.

Calvo, M. G., D. Beltrán, and A. Fernández-martín (2014). "Processing of facial expressions in peripheral vision : Neurophysiological evidence". In: *Biological Psychology* 100, pp. 60–70. DOI: 10.1016/j.biopsycho.2014.05.007.

Carlson, T. et al. (2013). "Reaction Time for Object Categorization Is Predicted by Representational Distance". In: *Journal of Cognitive Neuroscience* 26.1, pp. 132–142. DOI: 10.1162/jocn.

Carlson, T. et al. (2017). "Ghosts in machine learning for cognitive neuroscience: Moving from data to theory". In: *NeuroImage* 180, pp. 88–100. DOI: 10.1016/j.neuroimage.2017.08.019.

Carlson, T. A., P. Schrater, and S. He (2003). "Patterns of activity in the categorical representations of objects." In: *Journal of cognitive neuroscience* 15.5, pp. 704–17. DOI: 10.1162/089892903322307429.

Carretié, L. (2014). "Exogenous (automatic) attention to emotional stimuli: a review". In: *Cognitive, Affective and Behavioral Neuroscience* 14.4, pp. 1228–1258. DOI: 10.3758/s13415-014-0270-2.

Cauchoix, M. et al. (2012). "The neural dynamics of visual processing in monkey extrastriate cortex: A comparison between univariate and multivariate techniques". In: *Machine learning and interpretation in neuroimaging*. New York: Springer, pp. 164–171. DOI: 10.1007/978-3-642-34713-9-21.

Cauchoix, M. et al. (2014). "The neural dynamics of face detection in the wild revealed by MVPA." In: *Journal of Neuroscience* 34.3, pp. 846–54. DOI: 10.1523/JNEUROSCI.3030-13.2014.

Cauchoix, M. et al. (2016). "Fast ventral stream neural activity enables rapid visual categorization". In: *NeuroImage* 125, pp. 280–290. DOI: 10.1016/j.neuroimage.2015.10.012.

Cecotti, H et al. (2017). "Single-trial detection of event-related fields in MEG from the presentation of happy faces : Results of the Biomag 2016 data challenge". In: *Engineering in Medicine and Biology Society (EMBC), 2017 39th Annual International Conference of the IEEE*, pp. 4467–4470.

Chang, L. and D. Y. Tsao (2017). "The Code for Facial Identity in the Primate Brain." In: *Cell* 169.6, pp. 1013–1028. DOI: 10.1016/j.cell.2017.05.011.

Chen, C. et al. (2016). "The Attentional Dependence of Emotion Cognition Is Variable with the Competing Task". In: *Frontiers in Behavioral Neuroscience* 10.November, pp. 1–12. DOI: 10.3389/fnbeh.2016.00219.

Chikazoe, J. et al. (2014). "Population coding of affect across stimuli, modalities and individuals". In: *Nature Neuroscience* 17.8, pp. 1114–1122. DOI: 10.1038/nn.3749.

Cichy, R. M., F. M. Ramirez, and D. Pantazis (2015). "Can visual information encoded in cortical columns be decoded from magnetoencephalography data in humans ?" In: *NeuroImage* 121, pp. 193–204. DOI: 10.1016/j.neuroimage.2015.07.011.

Cichy, R. M. et al. (2016). "Dynamics of scene representations in the human brain revealed by magnetoencephalography and deep neural networks". In: *NeuroImage*. DOI: 10.1016/j.neuroimage.2016.03.063.

Clark, T. F. and D. N. Mcintosh (2008). "Autism and the Extraction of Emotion From Briefly Presented Facial Expressions : Stumbling at the First Step of Empathy". In: *Emotion* 8.6, pp. 803–809. DOI: 10.1037/a0014124.

Clarke, A. et al. (2013). "From perception to conception: How meaningful objects are processed over time". In: *Cerebral Cortex* 23.1, pp. 187–197. DOI: 10.1093/cercor/bhs002.

Claus, S. et al. (2012). "High frequency spectral components after Secobarbital: The contribution of muscular origin—A study with MEG/EEG". In: *Epilepsy Research* 100.1-2, pp. 132–141. DOI: 10.1016/j.eplepsyres.2012.02.002.

Cohen, D. (1968). "Magnetoencephalography: Evidence of Magnetic Fields Produced by Alpha-Rhythm Currents". In: *Science* 161.3843, pp. 784–786. DOI: 10.1126/science.161.3843.784.

Cohen, D (1972). "Magnetoencephalography: detection of the brain's electrical activity with a superconducting magnetometer." In: *Science* 175.4022, pp. 664–6.

Cohen, J. (1992). "A power primer". In: *Psychological Bulletin* 112.1, pp. 155–159. DOI: 10.1037/0033-2909.112.1.155.

Cohen, M. A. et al. (2017). "Visual search for object categories is predicted by the representational architecture of high-level visual cortex." In: *Journal of neurophysiology* 117.1, pp. 388–402. DOI: 10.1152/jn.00569.2016.

Combrisson, E. and K. Jerbi (2015). "Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy". In: *Journal of Neuroscience Methods* 250, pp. 126–136. DOI: 10.1016/j.jneumeth.2015.01.010.

Corbetta, M. and G. L. Shulman (2002). "Control of goal-directed and stimulus-driven attention in the brain". In: *Nature Reviews Neuroscience* 3.3, pp. 201–215. DOI: 10.1038/nrn755.

Cox, D. D. and R. L. Savoy (2003). "Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity

in human visual cortex". In: *NeuroImage* 19, pp. 261–270. DOI: 10.1016/S1053-8119(03)00049-1.

Cox, D. D. (2014). "Do we understand high-level vision?" In: *Current Opinion in Neurobiology* 25, pp. 187–193. DOI: 10.1016/j.conb.2014.01.016.

Crouzet, S. M. et al. (2012). "Animal Detection Precedes Access to Scene Category". In: *PLoS ONE* 7.12, pp. 1–9. DOI: 10.1371/journal.pone.0051471.

Davidesco, I. et al. (2014). "Exemplar Selectivity Reflects Perceptual Similarities in the Human Fusiform Cortex". In: *Cerebral Cortex* 24, pp. 1879–1893. DOI: 10.1093/cercor/bht038.

De Cesarei, A. et al. (2018). "Categorization Goals Modulate the Use of Natural Scene Statistics". In: *Journal of Cognitive Neuroscience*, pp. 1–17. DOI: 10.1162/jocn{_}a{_}01333.

Dehaene, S. (2016). "Decoding the Dynamics of Conscious Perception : The Temporal Generalization Method". In: *Micro-, Meso- and Macro-Dynamics of the Brain*. Ed. by B. G and C. Y. New York: Springer, pp. 85–97. DOI: 10.1007/978-3-319-28802-4.

Del Zotto, M. and A. J. Pegna (2015). "Processing of masked and unmasked emotional faces under different attentional conditions : an electrophysiological investigation". In: *Frontiers in Psychology* 6, p. 1691. DOI: 10.3389/fpsyg.2015.01691.

Delorme, A. and S. Makeig (2004). "EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis". In: *Journal of Neuroscience Methods* 134.1, pp. 9–21. DOI: 10.1016/j.jneumeth.2003.10.009.

Devue, C. and G. M. Grimshaw (2017). "Faces are special, but facial expressions aren't: Insights from an oculomotor capture paradigm". In: *Attention, Perception, and Psychophysics* 79.5, pp. 1438–1452. DOI: 10.3758/s13414-017-1313-x.

DeWit, L. et al. (2016). "Is neuroimaging measuring information in the brain ?" In: *Psychonomic Bulletin & Review* 23.5, pp. 1415–1428. DOI: 10.3758/s13423-016-1002-0.

Diamond, R and S Carey (1986). "Why faces are and are not special: an effect of expertise". In: *Journal of experimental psychology* 115.2, pp. 107–117.

DiCarlo, J. J., D. Zoccolan, and N. C. Rust (2012). "How does the brain solve visual object recognition?" In: *Neuron* 73.3, pp. 415–34. DOI: 10.1016/j.neuron.2012.01.010.

Diepen, R. M. van et al. (2016). "The Role of Alpha Activity in Spatial and Feature-Based Attention." In: *eNeuro* 3.5, 1–11. DOI: 10.1523/ENEURO.0204-16.2016.

Dilks, D. D. et al. (2013). "The Occipital Place Area Is Causally and Selectively Involved in Scene Perception". In: *Journal of Neuroscience* 33.4, pp. 1331–1336. DOI: 10.1523/JNEUROSCI.4081-12.2013.

Dima, D. C. and K. D. Singh (2018). "Dynamic representations of behaviourally relevant features support rapid face processing in the human ventral visual stream". In: *bioRxiv* August 2018. DOI: 10.1101/394916.

Dima, D. C., G. Perry, and K. D. Singh (2018a). "Spatial frequency supports the emergence of categorical representations in visual cortex during natural scene perception". In: *NeuroImage* 179, pp. 102–116. DOI: 10.1016/j.neuroimage.2018.06.033.

Dima, D. C. et al. (2018b). "Spatiotemporal dynamics in human visual cortex rapidly encode the emotional content of faces". In: *Human Brain Mapping* 39.10, pp. 3993–4006. DOI: 10.1002/hbm.24226.

Dimigen, O. et al. (2011). "Coregistration of eye movements and EEG in natural reading: Analyses and review". In: *Journal of Experimental Psychology: General* 140.4, pp. 552–572. DOI: 10.1037/a0023885.

Domingos, P. (2012). "A few useful things to know about machine learning". In: *Communications of the ACM* 55.10, p. 78. DOI: 10.1145/2347736.2347755.

Downing, P. E. and M. V. Peelen (2011). "The role of occipitotemporal body-selective regions in person perception". In: *Cognitive Neuroscience* 2.3-4, pp. 186–203. DOI: 10.1080/17588928.2011.582945.

Du, S. and A. M. Martinez (2013). "Wait, are you sad or angry? Large exposure time differences required for the categorization of facial expressions of emotion". In: *Journal of Vision* 13.4, pp. 13–13. DOI: 10.1167/13.4.13.

Duchaine, B. and G. Yovel (2015). "A revised neural framework for face processing". In: *Annual Review of Vision Science* 1, pp. 393–416. DOI: 10.1146/annurev-vision-082114-035518.

Edelman, S. et al. (1998). "Toward direct visualization of the internal shape represetation space by fMRI". In: *Psychobiology* 26.4, pp. 309–321. DOI: 10.1093/cercor/11.10.946.

Efron, B. and R. Tibshirani (1986). "Bootstrap Methods for Standard Errors , Confidence Intervals , and Other Measures of Statistical Accuracy". In: *Statistical Science* 1.1, pp. 54–75.

Efron, B. (1987). "Better Bootstrap Confidence Intervals". In: *Journal of the American Statistical Association* 82.397, pp. 171–185.

Eger, E. et al. (2003). "Rapid extraction of emotional expression: Evidence from evoked potential fields during brief presentation of face stimuli". In: *Neuropsychologia* 41.7, pp. 808–817. DOI: 10.1016/S0028-3932(02)00287-7.

Eimer, M (1996). "The N2pc component as an indicator of attentional selectivity." In: *Electroencephalography and clinical neurophysiology* 99.3, pp. 225–34.

Eimer, M, A Holmes, and F. P. McGlone (2003). "The role of spatial attention in the processing of facial expression: an ERP study of rapid brain responses to six basic emotions". In: *Cogn Affect Behav Neurosci* 3.2, pp. 97–110. DOI: 10.3758/CABN.3.2.97.

Eimer, M. (2000). "Attentional modulations of event-related brain potentials sensitive to faces". In: *Cognitive Neuropsychology* 17.1/2/3, pp. 103–116. DOI: 10.1080/026432900380517.

Eimer, M. et al. (2011). "The N170 component and its links to configural face processing: A rapid neural adaptation study". In: *Brain Research* 1376, pp. 76–87. DOI: 10.1016/J.BRAINRES.2010.12.046.

Ekman, P and W. Friesen (1977). *Facial action coding system: a technique for the measurement of facial movement*. Palo Alto, CA: Consulting Psychologists Press.

Engbert, R. and R. Kliegl (2003). "Microsaccades uncover the orientation of covert attention". In: *Vision Research* 43, pp. 1035–1045. DOI: 10.1016/S0042-6989(03)00084-1.

Epstein, R. A. (2008). "Parahippocampal and retrosplenial contributions to human spatial navigation". In: *Trends in Cognitive Sciences* 12.10, pp. 388–396. DOI: 10.1016/j.tics.2008.07.004.

Epstein, R. A. and N. Kanwisher (1998). "A cortical representation of the local visual environment." In: *Nature* 392, pp. 598–601. DOI: 10.1038/33402.

Erez, Y. and J. Duncan (2015). "Discrimination of Visual Categories Based on Behavioral Relevance in Widespread Regions of Frontoparietal Cortex". In: *Journal of Neuroscience* 35.36, pp. 12383–12393. DOI: 10.1523/JNEUROSCI.1134-15.2015.

Etzel, J. A., J. M. Zacks, and T. S. Braver (2013). "Searchlight analysis: Promise, pitfalls, and potential". In: *NeuroImage* 78, pp. 261–269. DOI: 10.1016/j.neuroimage.2013.03.041.

Ewbank, M. P. et al. (2009). "Anxiety predicts a differential neural response to attended and unattended facial signals of anger and fear". In: *NeuroImage* 44.3, pp. 1144–1151. DOI: 10.1016/J.NEUROIMAGE.2008.09.056.

Fairhall, A. (2014). "The receptive field is dead. Long live the receptive field?" In: *Current Opinion in Neurobiology* 25, pp. 9–12. DOI: 10.1016/j.conb.2014.02.001.

Fan, R.-E. et al. (2008). "LIBLINEAR: A Library for Large Linear Classification". In: *Journal of Machine Learning Research* 9.2008, pp. 1871–1874. DOI: 10.1038/oby.2011.351.

Farah, M. J., K. D. Wilson, and J. N. Tanaka (1998). "What Is " Special " About Face Perception ?" In: *Psychological review* 105.3, pp. 482–498. DOI: 10.1037//0033-295X.105.3.482.

Feldmann-Wüstefeld, T., M. Schmidt-Daffy, and A. Schubö (2011). "Neural evidence for the threat detection advantage: Differential attention allocation to angry and happy faces". In: *Psychophysiology* 48.5, pp. 697–707. DOI: 10.1111/j.1469-8986.2010.01130.x.

Felleman, D. J. and D. C. Van Essen (1991). "Distributed hierachical processing in the primate cerebral cortex". In: *Cerebral Cortex* 1.1, pp. 1–47. DOI: 10.1093/cercor/1.1.1.

Fenker, D. B. et al. (2010). "Mandatory Processing of Irrelevant Fearful Face Features in Visual Search". In: *Journal of Cognitive Neuroscience* 22.12, pp. 2926–2938. DOI: 10.1162/jocn.2009.21340.

Field, D. J. (1987). "Relations between the statistics of natural images and the response properties of cortical cells." In: *Journal of the Optical Society of America. A, Optics and image science* 4.12, pp. 2379–2394. DOI: 10.1364/JOSAA.4.002379.

Fisch, L. et al. (2010). "Neural "Ignition": Enhanced Activation Linked to Perceptual Awareness in Human Ventral Stream Visual Cortex". In: *Neuron* 64.4, pp. 562–574. DOI: 10.1016/j.neuron.2009.11.001.Neural.

Fisher, K., J. Towler, and M. Eimer (2016). "Facial identity and facial expression are initially integrated at visual perceptual stages of face processing". In: *Neuropsychologia* 80, pp. 115–125. DOI: 10.1016/j.neuropsychologia.2015.11.011.

Folstein, J. R., T. J. Palmeri, and I. Gauthier (2014). "Perceptual advantage for category-relevant perceptual dimensions: the case of shape and motion". In: *Frontiers in Psychology* 5, p. 1394. DOI: 10.3389/fpsyg.2014.01394.

Folstein, J. R. et al. (2015). "Category Learning Stretches Neural Representations in Visual Cortex". In: *Current Directions in Psychological Science* 24.1, pp. 17–23. DOI: 10.1177/0963721414550707.

Fox, E. and L. Damjanovic (2006). "The eyes are sufficient to produce a threat superiority effect." In: *Emotion* 6.3, pp. 534–9. DOI: 10.1037/1528-3542.6.3.534.

Fox, E. et al. (2000). "Facial Expressions of Emotion: Are Angry Faces Detected More Efficiently?" In: *Cognition & Emotion* 14.1, pp. 61–92. DOI: 10.1080/026999300378996.

Fox, E. et al. (2002). "Attentional bias for threat : Evidence for delayed disengagement from emotional faces". In: *Cognition and Emotion* 16.3, pp. 355–379. DOI: 10.1080/02699930143000527.

Fox, E., N. Derakshan, and L. Shoker (2008). "Trait anxiety modulates the electrophysiological indices of rapid spatial orienting towards angry faces". In: *NeuroReport* 19.3, pp. 259–263. DOI: 10.1097/WNR.0b013e3282f53d2a.

Freiwald, W., B. Duchaine, and G. Yovel (2016). "Face Processing Systems: From Neurons to Real-World Social Perception". In: *Annual Review of Neuroscience* 39.1, pp. 325–346. DOI: 10.1146/annurev-neuro-070815-013934.

Friston, K. J. (2009). "Modalities, Modes, and Models in Functional Neuroimaging". In: *Science* 326, pp. 399–403. DOI: 10.1126/science.1174521.

Frühholz, S., A. Jellinghaus, and M. Herrmann (2011). "Time course of implicit processing and explicit processing of emotional faces and emotional words". In: *Biological Psychology* 87.2, pp. 265–274. DOI: 10.1016/j.biopsycho.2011.03.008.

Furl, N., M. Lohse, and F. Pizzorni-Ferrarese (2017). "Low-frequency oscillations employ a general coding of the spatio-temporal similarity of dynamic faces". In: *NeuroImage* 157, pp. 486–499. DOI: 10.1016/j.neuroimage.2017.06.023.

Fusar-Poli, P. et al. (2009). "Functional atlas of emotional faces processing: A voxel-based meta-analysis of 105 functional magnetic resonance imaging studies". In: *Journal of Psychiatry and Neuroscience* 34.6, pp. 418–432. DOI: 10.1016/S1180-4882(09)50077-7.

García-Pérez, M. A. (2001). "Yes-No Staircases with Fixed Step Sizes : Psychometric Properties and Optimal Setup". In: *Optometry and Vision Science* 78.1, pp. 56–64.

Garrido, M. et al. (2012). "Functional Evidence for a Dual Route to Amygdala". In: *Current Biology* 22.2, pp. 129–134. DOI: 10.1016/J.CUB.2011.11.056.

Garvert, M. M. et al. (2014). "Subcortical amygdala pathways enable rapid face processing". In: *NeuroImage* 102.P2, pp. 309–316. DOI: 10.1016/j.neuroimage.2014.07.047.

Gerven, M. V. et al. (2009). "Selecting features for BCI control based on a covert spatial attention paradigm". In: *Neural Networks* 22.9, pp. 1271–1277. DOI: 10.1016/j.neunet.2009.06.004.

Gilbert, C. D. and W. Li (2013). "Top-down influences on visual processing". In: *Nature Reviews Neuroscience* 14.5, pp. 350–363. DOI: 10.1038/nrn3476.

Gilbert, C. D. and M. Sigman (2007). "Brain States: Top-Down Influences in Sensory Processing". In: *Neuron* 54.5, pp. 677–696. DOI: 10.1016/J.NEURON.2007.05.019.

Gitelman, D. R. et al. (1999). "A large-scale distributed network for covert spatial attention. Further anatomical delineation based on stringent behavioural and cognitive controls". In: *Brain* 122.6, pp. 1093–1106. DOI: 10.1093/brain/122.6.1093.

Goddard, E. et al. (2017). "Interpreting the dimensions of neural feature representations revealed by dimensionality reduction". In: *NeuroImage* 180, pp. 41–67. DOI: 10.1016/j.neuroimage.2017.06.068.

Gohel, B. et al. (2018). "Dynamic pattern decoding of source-reconstructed MEG or EEG data: Perspective of multivariate pattern analysis and signal leakage". In: *Computers in Biology and Medicine* 93, pp. 106–116. DOI: 10.1016/j.compbiomed.2017.12.020.

Golarai, G. et al. (2015). "Distinct representations of configural and part information across multiple face-selective regions of the human brain." In: *Frontiers in Psychology* 6, p. 1710. DOI: 10.3389/fpsyg.2015.01710.

Goodale, M. A. and A. D. Milner (1992). "Separate Visual Pathways for Perception and Action". In: *Trends in Neurosciences* 15.1, pp. 20–25. DOI: 10.1016/0166-2236(92)90344-8.

Gray, K. L. H. et al. (2013). "Faces and awareness: low-level, not emotional factors determine perceptual dominance." In: *Emotion* 13.3, pp. 537–44. DOI: 10.1037/a0031403.

Greene, M. R. and B. C. Hansen (2018). "Shared spatiotemporal category representations in biological and artificial deep neural networks". In: *PLOS Computational Biology* 14.7, e1006327. DOI: 10.1371/journal.pcbi.1006327.

Greene, M. R. et al. (2016). "Visual Scenes are Categorized by Function". In: *Journal of Experimental Psychology: General* 145.1, pp. 82–94. DOI: 10.1037/xge0000129.Visual.

Grill-Spector, K. and K. S. Weiner (2014). "The functional architecture of the ventral temporal cortex and its role in categorization". In: *Nature Reviews Neuroscience* 15.8, pp. 536–548. DOI: 10.1038/nrn3747.

Grill-Spector, K. et al. (2018). "The functional neuroanatomy of face perception: From brain measurements to deep neural networks". In: *Interface Focus* 8. DOI: 10.1098/rsfs.2018.0013.

Groen, I. I. A. et al. (2013). "From Image Statistics to Scene Gist: Evoked Neural Activity Reveals Transition from Low-Level Natural Image Structure to Scene Category". In: *Journal of Neuroscience* 33.48, pp. 18814–18824. DOI: 10.1523/JNEUROSCI.3128-13.2013.

Groen, I. I. A. et al. (2016). "The time course of natural scene perception with reduced attention". In: *Journal of Neurophysiology* 115.2, pp. 931–946. DOI: 10.1152/jn.00896.2015.

Groen, I. I. A., E. H. Silson, and C. I. Baker (2017). "Contributions of low- and high-level properties to neural processing of visual scenes in the human brain". In: *Philosophical transactions of the Royal Society of London. Series B* 372. DOI: http://dx.doi.org/10.1098/rstb.2016.0102.

Groen, I. I. et al. (2018). "Distinct contributions of functional and deep neural network features to representational similarity of scenes in human brain and behavior". In: *eLife* 7, e32962. DOI: 10.7554/eLife.32962.

Grootswagers, T., S. G. Wardle, and T. A. Carlson (2017). "Decoding Dynamic Brain Patterns from Evoked Responses: A Tutorial on Multivariate Pattern Analysis Applied to Time Series Neuroimaging Data". In: *Journal of Cognitive Neuroscience* 29.4, pp. 677–697. DOI: 10.1162/jocn{_}a{_}01068.

Grootswagers, T., R. M. Cichy, and T. A. Carlson (2018). "Finding decodable information that can be read out in behaviour". In: *NeuroImage* 179, pp. 252–262. DOI: 10.1016/J.NEUROIMAGE.2018.06.022.

Gross, C. G. (2002). "Genealogy of the "Grandmother Cell"". In: *The Neuroscientist* 8.5, pp. 512–518. DOI: 10.1177/107385802237175.

Güçlü, U. and M. A. J. van Gerven (2014). "Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Brain's Ventral Visual Pathway". In: 35.27, pp. 10005–10014. DOI: 10.1523/JNEUROSCI.5023-14.2015.

Guggenmos, M., P. Sterzer, and R. M. Cichy (2018). "Multivariate pattern analysis for MEG: A comparison of dissimilarity measures". In: *NeuroImage* 173, pp. 434–447. DOI: 10.1016/J.NEUROIMAGE.2018.02.044.

Guyader, N. et al. (2004). "Image phase or amplitude? Rapid scene categorization is an amplitude-based process". In: *Comptes Rendus - Biologies* 327, pp. 313–318. DOI: 10.1016/j.crvi.2004.02.006.

Halgren, E. et al. (2000). "Cognitive Response Profile of the Human Fusiform Face Area as Determined by MEG". In: *Cerebral Cortex* 10, pp. 69–81.

Hämäläinen, M. et al. (1993). "Magnetoencephalography theory, instrumentation, and applications to noninvasive studies of the working human brain". In: *Reviews of Modern Physics* 65.2, pp. 413–497. DOI: 10.1103/RevModPhys.65.413.

Harel, A. et al. (2016). "The Temporal Dynamics of Scene Processing: A Multi-faceted EEG Investigation". In: *eNeuro* 3.5, pp. 1–18. DOI: 10.1523/ENEURO.0139-16.2016.

Hari, R. and R. Salmelin (2012). "Magnetoencephalography: From SQUIDs to neuroscience. Neuroimage 20th Anniversary Special Edition." In: *NeuroImage* 61.2, pp. 386–396. DOI: 10.1016/j.neuroimage.2011.11.074.

Hassabis, D. et al. (2017). "Neuroscience-Inspired Artificial Intelligence". In: *Neuron* 95.2, pp. 245–258. DOI: 10.1016/j.neuron.2017.06.011.

Hasselmo, M. E., E. T. Rolls, and G. C. Baylis (1989). "The role of expression and identity in the face-selective responses of neurons in the temporal visual cortex of the monkey". In: *Behavioural Brain Research* 32.3, pp. 203–218. DOI: 10.1016/S0166-4328(89)80054-3.

Haufe, S. et al. (2014). "On the interpretation of weight vectors of linear models in multivariate neuroimaging". In: *NeuroImage* 87, pp. 96–110. DOI: 10.1016/j.neuroimage.2013.10.067.

Haxby, J. V. et al. (2001). "Distributed and overlapping representations of faces and objects in ventral temporal cortex." In: *Science* 293.5539, pp. 2425–30. DOI: 10.1126/science.1063736.

Haxby, J. V., E. A. Hoffman, and M. I. Gobbini (2000). "The distributed human neural system for face perception". In: *Trends in Cognitive Sciences* 4.6, pp. 223–233.

Haxby, J. V. et al. (2011). "A common, high-dimensional model of the representational space in human ventral temporal cortex". In: *Neuron* 72.2, pp. 404–416. DOI: 10.1016/j.neuron.2011.08.026.

Haxby, J. V., A. C. Connolly, and J. S. Guntupalli (2014). "Decoding Neural Representational Spaces Using Multivariate Pattern Analysis". In: *Annual Review of Neuroscience* 37.1, pp. 435–456. DOI: 10.1146/annurev-neuro-062012-170325.

Haynes, J.-D. (2015). "A Primer on Pattern-Based Approaches to fMRI: Principles, Pitfalls, and Perspectives". In: *Neuron* 87.2, pp. 257–270. DOI: 10.1016/j.neuron.2015.05.025.

Haynes, J.-D. and G. Rees (2006). "Decoding mental states from brain activity in humans". In: *Nature Reviews Neuroscience* 7, pp. 523–534. DOI: 10.1038/nrn1931.

Hebart, M. N. and C. I. Baker (2017). "Deconstructing multivariate decoding for the study of brain function". In: *NeuroImage* July, pp. 1–15. DOI: 10.1016/j.neuroimage.2017.08.005.

Hebart, M. N. et al. (2018). "The representational dynamics of task and object processing in humans". In: *eLife* 7, e32816. DOI: 10.7554/eLife.32816.

Hedger, N., W. J. Adams, and M. Garner (2015). "Fearful faces have a sensory advantage in the competition for awareness". In: *Journal of Experimental Psychology: Human Perception and Performance* 41.6, pp. 1748–1757. DOI: 10.1037/xhp0000127.

Hedger, N. et al. (2016). "Are visual threats prioritised without awareness? A critical review and meta analysis involving 3 behavioural paradigms and 2696 observers." In: *Psychological Bulletin* 142.9, pp. 934–968.

Henriksson, L., M. Mur, and N. Kriegeskorte (2015). "Faciotopy-A face-feature map with face-like topology in the human occipital face area". In: *Cortex* 72, pp. 156–167. DOI: 10.1016/j.cortex.2015.06.030.

Herrmann, M. J. et al. (2008). "Enhancement of activity of the primary visual cortex during processing of emotional stimuli as measured with event-related functional near-infrared spectroscopy and event-related potentials." In: *Human brain mapping* 29.1, pp. 28–35. DOI: 10.1002/hbm.20368.

Hillebrand, A and G. R. Barnes (2002). "A Quantitative Assessment of the Sensitivity of Whole-Head MEG to Activity in the Adult Human Cortex". In: *NeuroImage* 16, pp. 638–650. DOI: 10.1006/nimg.2002.1102.

Hillebrand, A. et al. (2005). "A new approach to neuroimaging with magnetoencephalography." In: *Human Brain Mapping* 25.2, pp. 199–211. DOI: 10.1002/hbm.20102.

Hillebrand, A. et al. (2012). "Frequency-dependent functional connectivity within resting-state networks : An atlas-based MEG beamformer solution". In: *NeuroImage* 59.4, pp. 3909–3921. DOI: 10.1016/j.neuroimage.2011.11.005.

Hinojosa, J. A., F Mercado, and L Carretié (2015). "N170 sensitivity to facial expression: A meta-analysis." In: *Neuroscience and Biobehavioral Reviews* 55, pp. 498–509. DOI: 10.1016/j.neubiorev.2015.06.002.

Hodsoll, S., E. Viding, and N. Lavie (2011). "Attentional capture by irrelevant emotional distractor faces." In: *Emotion* 11.2, pp. 346–353. DOI: 10.1037/a0022771.

Holmes, A., P. Vuilleumier, and M. Eimer (2003). "The processing of emotional facial expression is gated by spatial attention: Evidence from event-related brain potentials". In: *Cognitive Brain Research* 16.2, pp. 174–184. DOI: 10.1016/S0926-6410(02)00268-9.

Holmes, A., J. S. Winston, and M. Eimer (2005). "The role of spatial frequency information for ERP components sensitive to faces and emotional facial expression". In: *Cognitive Brain Research* 25.2, pp. 508–520. DOI: 10.1016/j.cogbrainres.2005.08.003.

Hong, H. et al. (2016). "Explicit information for category-orthogonal object properties increases along the ventral stream". In: *Nature Neuroscience* 19.4, pp. 613–622. DOI: 10.1038/nn.4247.

Horschig, J. M. et al. (2015). "Modulation of Posterior Alpha Activity by Spatial Attention Allows for Controlling A Continuous Brain–Computer Interface". In: *Brain Topography* 28.6, pp. 852–864. DOI: 10.1007/s10548-014-0401-7.

Howe, P. D. L. (2017). "Natural scenes can be identified as rapidly as individual features". In: *Atten Percept Psychophys* 79, pp. 1674–1681. DOI: 10.3758/s13414-017-1349-y.

Huang, S.-L., Y.-C. Chang, and Y.-J. Chen (2011). "Task-irrelevant angry faces capture attention in visual search while modulated by resources." In: *Emotion* 11.3, pp. 544–552. DOI: 10.1037/a0022763.

Hubel, D. H. and T. N. Wiesel (1962). "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex". In: *The Journal of Physiology* 160.1, pp. 106–154. DOI: 10.1113/jphysiol.1962.sp006837.

Hung, C. P. et al. (2005). "Fast Readout of Object Identify from Macaque Inferior Temporal Cortex". In: *Science* 310.5749, p. 4. DOI: 10.1126/science.1117593.

Iivanainen, J., M. Stenroos, and L. Parkkonen (2017). "Measuring MEG closer to the brain: Performance of on-scalp sensor arrays". In: *NeuroImage* 147, pp. 542–553. DOI: 10.1016/j.neuroimage.2016.12.048.

Ikeda, K., A. Sugiura, and T. Hasegawa (2013). "Fearful faces grab attention in the absence of late affective cortical responses". In: *Psychophysiology* 50, pp. 60–69. DOI: 10.1111/j.1469-8986.2012.01478.x.

Inuggi, A. et al. (2014). "Cortical response of the ventral attention network to unattended angry facial expressions : an EEG source analysis study". In: *Frontiers in Psychology* 5, p. 1498. DOI: 10.3389/fpsyg.2014.01498.

Ishai, A. et al. (2004). "Repetition suppression of faces is modulated by emotion". In: *Proceedings of the National Academy of Sciences* 101.26, pp. 9827–9832. DOI: 10.1073/pnas.0403559101.

Ishai, A. (2008). "Let's face it: It's a cortical network". In: *NeuroImage* 40.2, pp. 415–419. DOI: 10.1016/j.neuroimage.2007.10.040.

Ishai, A. et al. (1999). "The Representation of Objects in the Human Occipital and Temporal Cortex". In: *Journal of Cognitive Neuroscience* 12.Supplement 2, pp. 35–51.

Ishai, A., P. C. Bikle, and L. G. Ungerleider (2006). "Temporal dynamics of face repetition suppression". In: *Brain Research Bulletin* 70.4-6, pp. 289–295. DOI: 10.1016/j.brainresbull.2006.06.002.

Isik, L. et al. (2014). "The dynamics of invariant object recognition in the human visual system". In: *Journal of Neurophysiology* 111, pp. 91–102. DOI: 10.1152/jn.00394.2013.

Jamalabadi, H. et al. (2016). "Classification Based Hypothesis Testing in Neuroscience : Below-Chance Level Classification Rates and Overlooked Statistical Properties of Linear Parametric Classifiers". In: *Human Brain Mapping* 37, pp. 1842–1855. DOI: 10.1002/hbm.23140.

Jia, Y. et al. (2014). "Caffe: Convolutional Architecture for Fast Feature Embedding". In: *Proceedings of the ACM International Conference on Multimedia - MM '14*, pp. 675–678. DOI: 10.1145/2647868.2654889.

Jiang, Y. and S. He (2006). "Cortical Responses to Invisible Faces: Dissociating Subsystems for Facial-Information Processing". In: *Current Biology* 16.20, pp. 2023–2029. DOI: 10.1016/j.cub.2006.08.084.

Jiang, Y. et al. (2009). "Dynamics of processing invisible faces in the brain: Automatic neural encoding of facial expression information". In: *NeuroImage* 44.3, pp. 1171–1177. DOI: 10.1016/j.neuroimage.2008.09.038.Dynamics.

Jonas, E. and K. P. Kording (2017). "Could a Neuroscientist Understand a Microprocessor?" In: *PLOS Computational Biology* 13.1, e1005268. DOI: 10.1371/journal.pcbi.1005268.

Joubert, O. R. et al. (2007). "Processing scene context: Fast categorization and object interference". In: *Vision Research* 47.26, pp. 3286–3297. DOI: 10.1016/j.visres.2007.09.013.

Jozwik, K. M., N. Kriegeskorte, and M. Mur (2016). "Visual features as stepping stones toward semantics : Explaining object similarity in IT and perception with non-negative least squares". In: *Neuropsychologia* 83, pp. 201–226. DOI: 10.1016/j.neuropsychologia.2015.10.023.

Kaiser, D., D. C. Azzalini, and M. V. Peelen (2016). "Shape-independent object category responses revealed by MEG and fMRI decoding". In: *Journal of Neurophysiology* 115, pp. 2246–2250. DOI: 10.1152/jn.01074.2015.

Kamitani, Y. and F. Tong (2005). "Decoding the visual and subjective contents of the human brain". In: *Nature Neuroscience* 8.5, pp. 679–685. DOI: 10.1038/nn1444.

Kanwisher, N, J McDermott, and M. M. Chun (1997). "The fusiform face area: a module in human extrastriate cortex specialized for face perception." In: *The Journal of Neuroscience* 17.11, pp. 4302–11. DOI: 10.1098/Rstb.2006.1934.

Kanwisher, N. (2000). "Domain specificity in face perception". In: *Nature Neuroscience* 3.8, pp. 759–763. DOI: 10.1038/77664.

Kauffmann, L., S. Ramanoël, and C. Peyrin (2014). "The neural bases of spatial frequency processing during scene perception." In: *Frontiers in Integrative Neuroscience* 8, p. 37. DOI: 10.3389/fnint.2014.00037.

Kauffmann, L. et al. (2015a). "Rapid scene categorization: Role of spatial frequency order, accumulation mode and luminance contrast". In: *Vision Research* 107, pp. 49–57. DOI: 10.1016/j.visres.2014.11.013.

Kauffmann, L. et al. (2015b). "Spatial frequency processing in scene-selective cortical regions". In: *NeuroImage* 112, pp. 86–95. DOI: 10.1016/j.neuroimage.2015.02.058.

Kauffmann, L. et al. (2017). "How does information from low and high spatial frequencies interact during scene categorization ?" In: *Visual Cognition* 25.9-10, pp. 853–867. DOI: 10.1080/13506285.2017.1347590.

Kelly, S. P. et al. (2005). "Visual Spatial Attention Control in an Independent Brain-Computer Interface". In: *IEEE Transactions on Biomedical Engineering* 52.9, pp. 1588–1596. DOI: 10.1109/TBME.2005.851510.

Kemp, R. I. et al. (2016). "Improving Unfamiliar Face Matching by Masking the External Facial Features". In: *Applied Cognitive Psychology* 30.4, pp. 622–627. DOI: 10.1002/acp.3239.

Khaligh-Razavi, S. M. and N. Kriegeskorte (2014). "Deep Supervised, but Not Unsupervised, Models May Explain IT Cortical Representation". In: *PLoS Computational Biology* 10.11, e1003915. DOI: 10.1371/journal.pcbi.1003915.

Kietzmann, T. C., P. Mcclure, and N. Kriegeskorte (2017). "Deep Neural Networks in Computational Neuroscience". In: *bioRxiv* May 2017. DOI: 10.1101/133504.

King, J.-R. and S Dehaene (2014). "Characterizing the dynamics of mental representations : the temporal generalization method". In: *Trends in Cognitive Sciences* 18.4, pp. 203–210. DOI: 10.1016/j.tics.2014.01.002.

Kleiner, M. et al. (2007). "What's new in Psychtoolbox-3?" In: *Perception* 36, S14. DOI: 10.1068/v070821.

Kliemann, D. et al. (2016). "Decoding task and stimulus representations in face-responsive cortex". In: *Cognitive Neuropsychology* 33.7-8, pp. 362–377. DOI: 10.1080/02643294.2016.1256873.

Knösche, T. R. (2002). "Transformation of whole-head MEG recordings between different sensor positions." In: *Biomedical Engineering* 47.3, pp. 59–62. DOI: 10.1515/bmte.2002.47.3.59.

Koelewijn, L. et al. (2013). "Spatial attention increases high-frequency gamma synchronisation in human medial visual cortex". In: *NeuroImage* 79, pp. 295–303. DOI: 10.1016/j.neuroimage.2013.04.108.

Kohler, C. G. et al. (2011). "Facial emotion perception in depression and bipolar disorder: A quantitative review". In: *Psychiatry Research* 188.3, pp. 303–309. DOI: 10.1016/j.psychres.2011.04.019.

Kok, P., J. F. M. Jehee, and F. P. de Lange (2012). "Less is more: expectation sharpens representations in the primary visual cortex." In: *Neuron* 75.2, pp. 265–70. DOI: 10.1016/j.neuron.2012.04.034.

Koster, E. H. W. et al. (2007). "Attention for Emotional Faces Under Restricted Awareness Revisited : Do Emotional Faces Automatically Attract Attention ?" In: *Emotion* 7.2, pp. 285–295. DOI: 10.1037/1528-3542.7.2.285.

Kragel, P. A. et al. (2018). "Representation, Pattern Information, and Brain Signatures: From Neurons to Neuroimaging". In: *Neuron* 99.2, pp. 257–273. DOI: 10.1016/j.neuron.2018.06.009.

Kravitz, D. J., C. S. Peng, and C. I. Baker (2011). "Real-World Scene Representations in High-Level Visual Cortex: It's the Spaces More Than the Places". In: *The Journal of Neuroscience* 31.20, pp. 7322–7333. DOI: 10.1523/JNEUROSCI.4588-10.2011.

Kravitz, D. J. et al. (2013). "The ventral visual pathway: An expanded neural framework for the processing of object quality". In: *Trends in Cognitive Sciences* 17.1, pp. 26–49. DOI: 10.1016/j.tics.2012.10.011.

Kriegeskorte, N. (2011). "Pattern-information analysis : From stimulus decoding to computational-model testing". In: *NeuroImage* 56.2, pp. 411–421. DOI: 10.1016/j.neuroimage.2011.01.061.

Kriegeskorte, N. and R. A. Kievit (2013). "Representational geometry : integrating cognition , computation , and the brain". In: *Trends in Cognitive Sciences* 17.8, pp. 401–412. DOI: 10.1016/j.tics.2013.06.007.

Kriegeskorte, N., R. Goebel, and P. Bandettini (2006). "Information-based functional brain mapping". In: *Proceedings of the National Academy of Sciences* 103.10, pp. 3863–3868.

Kriegeskorte, N. et al. (2007). "Individual faces elicit distinct response patterns in human anterior temporal cortex." In: *Proceedings of the National Academy of Sciences of the United States of America* 104.51, pp. 20600–5. DOI: 10.1073/pnas.0705654104.

Kriegeskorte, N., M. Mur, and P. Bandettini (2008). "Representational similarity analysis - connecting the branches of systems neuroscience." In: *Frontiers in systems neuroscience* 2.November, p. 4. DOI: 10.3389/neuro.06.004.2008.

Krizhevsky, A., I. Sutskever, and G. E. Hinton (2012). "ImageNet Classification with Deep Convolutional Neural Networks". In: *Advances In Neural Information Processing Systems*, pp. 1–9. DOI: http://dx.doi.org/10.1016/j.protcy.2014.09.007.

Krolak-Salmon, P et al. (2001). "Processing of facial emotion expression: Spatio-temporal data as assessed by scalp event-related potentials". In: *European Journal of Neuroscience* 13, pp. 987–994. DOI: 10.1046/j.0953-816X.2001.01454.x.

Krolak-Salmon, P. et al. (2004). "Early amygdala reaction to fear spreading in occipital, temporal, and frontal cortex: A depth electrode ERP study in human". In: *Neuron* 42.4, pp. 665–676. DOI: 10.1016/S0896-6273(04)00264-8.

Lakens, D. (2017). "Equivalence Tests: A Practical Primer for t Tests, Correlations, and Meta-Analyses". In: *Social Psychological and Personality Science* 8.4, pp. 355–362. DOI: 10.1177/1948550617697177.

Lamme, V. A. F. and P. R. Roelfsema (2000). "The distinct modes of vision offered by feedforward and recurrent processing". In: *Trends in Neurosciences* 23.11, pp. 571–579. DOI: 10.1016/S0166-2236(00)01657-X.

Lamme, V. A. F., K. Zipser, and H. Spekreijse (2002). "Masking Interrupts Figure-Ground Signals in V1". In: *Journal of Cognitive Neuroscience* 14.7, pp. 1044–1053. DOI: 10.1162/089892902320474490.

Lange, K. et al. (2003). "Task instructions modulate neural responses to fearful facial expressions". In: *Biological Psychiatry* 53.3, pp. 226–232. DOI: 10.1016/S0006-3223(02)01455-5.

Lau, H. C. (2008). "Are we studying consciousness yet?" In: *Frontiers of Consciousness*. Oxford University Press, pp. 245–258. DOI: 10.1093/acprof:oso/9780199233151.003.0008.

Lavie, N. (2005). "Distracted and confused?: Selective attention under load". In: *Trends in Cognitive Sciences* 9.2, pp. 75–82. DOI: 10.1016/j.tics.2004.12.004.

LeDoux, J. E. and R. Brown (2017). "A higher-order theory of emotional consciousness". In: *Proceedings of the National Academy of Sciences* 114.10, E2016–E2025. DOI: 10.1073/pnas.1619316114.

Lemm, S. et al. (2011). "Introduction to machine learning for brain imaging". In: *NeuroImage* 56.2, pp. 387–399. DOI: 10.1016/j.neuroimage.2010.11.004.

Leopold, D. A. et al. (2001). "Prototype-referenced shape encoding revealed by high-level aftereffects". In: *Nature Neuroscience* 4.1, pp. 89–94. DOI: 10.1038/82947.

Lim, S.-L., S Padmala, and L Pessoa (2009). "Segregating the significant from the mundane on a moment-to-moment basis via direct and indirect amygdala contributions". In: *Proceedings of the National Academy of Sciences* 106.39, pp. 16841–16846. DOI: 10.1073/pnas.0904551106.

Liu, H. et al. (2009). "Timing, Timing, Timing: Fast Decoding of Object Information from Intracranial Field Potentials in Human Visual Cortex". In: *Neuron* 62.2, pp. 281–290. DOI: 10.1016/j.neuron.2009.02.025.

Liu, L. and A. A. Ioannides (2010). "Emotion separation is completed early and it depends on visual field presentation". In: *PLoS ONE* 5.3, e9790. DOI: 10.1371/journal.pone.0009790.

Lohse, M. et al. (2016). "Effective Connectivity from Early Visual Cortex to Posterior Occipitotemporal Face Areas Supports Face Selectivity and Predicts Developmental Prosopagnosia". In: *The Journal of neuroscience* 36.13, pp. 3821–3828. DOI: 10.1523/JNEUROSCI.3621-15.2016.

Long, B., C.-P. Yu, and T. Konkle (2018). "Mid-level visual features underlie the high-level categorical organization of the ventral stream". In: *Proceedings of the National Academy of Sciences* 115.38, :E9015–E9024. DOI: 10.1073/pnas.1719616115.

Longmore, C. A., C. H. Liu, and A. W. Young (2015). "The importance of internal facial features in learning new faces". In: *Quarterly Journal of Experimental Psychology* 68.2, pp. 249–260. DOI: 10.1080/17470218.2014.939666.

Lu, Q. et al. (2013). "Predicting depression based on dynamic regional connectivity: A windowed Granger causality analysis of MEG recordings". In: *Brain Research* 1535, pp. 52–60. DOI: 10.1016/j.brainres.2013.08.033.

Ly, A. et al. (2018). "Bayesian Reanalyses From Summary Statistics: A Guide for Academic Consumers". In: *Advances in Methods and Practices in Psychological Science* 1.3, pp. 367–374. DOI: 10.1177/2515245918779348.

Magazzini, L. and K. D. Singh (2017). "Spatial attention modulates visual gamma oscillations across the human ventral stream". In: *NeuroImage* 166, pp. 219–229. DOI: 10.1016/j.neuroimage.2017.10.069.

Maguire, J. F. and P. D. L. Howe (2016). "Failure to detect meaning in RSVP at 27 ms per picture". In: *Attention, Perception, & Psychophysics* 78.5, pp. 1405–1413. DOI: 10.3758/s13414-016-1096-5.

Marblestone, A. H., G. Wayne, and K. P. Kording (2016). "Toward an Integration of Deep Learning and Neuroscience". In: *Frontiers in Computational Neuroscience* 10, p. 94. DOI: 10.3389/fncom.2016.00094.

Marr, D. (1982). *Vision : a computational investigation into the human representation and processing of visual information*. San Francisco: W.H. Freeman, p. 397.

Maurer, D., R. L. Grand, and C. J. Mondloch (2002). "The many faces of configural processing". In: *Trends in Cognitive Sciences* 6.6, pp. 255–260. DOI: 10.1016/S1364-6613(02)01903-4.

McFadyen, J. et al. (2017). "A Rapid Subcortical Amygdala Route for Faces Irrespective of Spatial Frequency and Emotion". In: *The Journal of Neuroscience* 37.14, pp. 3864–3874. DOI: 10.1523/JNEUROSCI.3525-16.2017.

Méndez-Bértolo, C. et al. (2016). "A fast pathway for fear in human amygdala". In: *Nature Neuroscience* 19.8, pp. 1041–1049. DOI: 10.1038/nn.4324.

Mercure, E., F. Dick, and M. H. Johnson (2008). "Featural and configural face processing differentially modulate ERP components". In: *Brain Research* 1239, pp. 162–170. DOI: 10.1016/J.BRAINRES.2008.07.098.

Meyer, S. S. et al. (2017). "Using generative models to make probabilistic statements about hippocampal engagement in MEG". In: *NeuroImage* 149, pp. 468–482. DOI: 10.1016/j.neuroimage.2017.01.029.

Micallef, L. and P. Rodgers (2014). "eulerAPE: Drawing Area-Proportional 3-Venn Diagrams Using Ellipses". In: *PLoS ONE* 9.7, e101717. DOI: 10.1371/journal.pone.0101717.

Mohanty, A. and T. J. Sussman (2013). "Top-down modulation of attention by emotion". In: *Frontiers in Human Neuroscience* 7, p. 102. DOI: 10.3389/fnhum.2013.00102.

Mohanty, A. et al. (2009). "Search for a threatening target triggers limbic guidance of spatial attention." In: *Journal of neuroscience* 29.34, pp. 10563–72. DOI: 10.1523/JNEUROSCI.1170-09.2009.

Mohr, S. et al. (2018). "Early identity recognition of familiar faces is not dependent on holistic processing". In: *bioRxiv* March 2018.

Mohsenzadeh, Y. et al. (2018a). "The Perceptual Neural Trace of Memorable Unseen Scenes". In: *bioRxiv* September 2018. DOI: 10.1101/414052.

Mohsenzadeh, Y. et al. (2018b). "Ultra-Rapid serial visual presentation reveals dynamics of feedforward and feedback processes in the ventral visual pathway". In: *eLife* 7.e36329. DOI: 10.7554/eLife.36329.001.

Moors, A. and J. De Houwer (2006). "Automaticity: A Theoretical and Conceptual Analysis." In: *Psychological Bulletin* 132.2, pp. 297–326. DOI: 10.1037/0033-2909.132.2.297.

Morawetz, C. et al. (2010). "Diverting Attention Suppresses Human Amygdala Responses to Faces". In: *Frontiers in Human Neuroscience* 4, p. 226. DOI: 10.3389/FNHUM.2010.00226.

Morel, S. et al. (2009). "EEG-MEG evidence for early differential repetition effects for fearful, happy and neutral faces". In: *Brain Research* 1254, pp. 84–98. DOI: 10.1016/j.brainres.2008.11.079.

Morris, J. S., A Ohman, and R. J. Dolan (1998). "Conscious and unconscious emotional learning in the human amygdala". In: *Nature* 393, pp. 467–470.

Mosher, J. C., R. M. Leahy, and P. S. Lewis (1999). "EEG and MEG: Forward solutions for inverse methods". In: *IEEE Transactions on Biomedical Engineering* 46.3, pp. 245–259. DOI: 10.1109/10.748978.

Mothes-Lasch, M. et al. (2011). "Visual Attention Modulates Brain Activation to Angry Voices". In: *Journal of Neuroscience* 31.26, pp. 9594–9598. DOI: 10.1523/JNEUROSCI.6665-10.2011.

Müsch, K. et al. (2016). "Gamma-band activity reflects attentional guidance by facial expression". In: *NeuroImage* 146, pp. 1142–1148. DOI: 10.1016/j.neuroimage.2016.09.025.

Muthukumaraswamy, S. D. (2013). "High-frequency brain activity and muscle artifacts in MEG/EEG: a review and recommendations." In: *Frontiers in human neuroscience* 7, p. 138. DOI: 10.3389/fnhum.2013.00138.

Muthukumaraswamy, S. D. and K. D. Singh (2008). "Spatiotemporal frequency tuning of BOLD and gamma band MEG responses compared in primary visual cortex". In: *NeuroImage* 40.4, pp. 1552–1560. DOI: 10.1016/j.neuroimage.2008.01.052.

Namdar, G., G. Avidan, and T. Ganel (2015). "Effects of configural processing on the perceptual spatial resolution for face features". In: *Cortex* 72, pp. 115–123. DOI: 10.1016/j.cortex.2015.04.007.

Naselaris, T. (2015). "Resolving Ambiguities of MVPA Using Explicit Models of Representation". In: *Trends in Cognitive Sciences* 19.10, pp. 551–554. DOI: 10.1016/j.tics.2015.07.005.

Nasr, S. and R. B. H. Tootell (2012). "A Cardinal Orientation Bias in Scene-Selective Visual Cortex". In: *Journal of Neuroscience* 32.43, pp. 14921–14926. DOI: 10.1523/JNEUROSCI.2036-12.2012.

Nasr, S. et al. (2011). "Scene-Selective Cortical Regions in Human and Nonhuman Primates". In: *Journal of Neuroscience* 31.39, pp. 13771–13785. DOI: 10.1523/JNEUROSCI.2792-11.2011.

Nasr, S., C. E. Echavarria, and R. B. H. Tootell (2014). "Thinking Outside the Box: Rectilinear Shapes Selectively Activate Scene-Selective Cortex". In: *Journal of Neuroscience* 34.20, pp. 6721–6735. DOI: 10.1523/JNEUROSCI.4802-13.2014.

Nastase, S. A. et al. (2017). "Attention selectively reshapes the geometry of distributed semantic representation". In: *Cerebral Cortex* 27.8, pp. 4277–4291. DOI: 10.1093/cercor/bhx138.

Navajas, J., M. Ahmadi, and R. Quian Quiroga (2013). "Uncovering the Mechanisms of Conscious Face Perception: A Single-Trial Study of the N170 Responses". In: *Journal of Neuroscience* 33.4, pp. 1337–1343. DOI: 10.1523/JNEUROSCI.1226-12.2013.

Nemrodov, D. et al. (2016). "The time course of individual face recognition : A pattern analysis of ERP signals". In: *NeuroImage* 132, pp. 469–476. DOI: 10.1016/j.neuroimage.2016.03.006.

Newen, A. and P. Vetter (2017). "Why cognitive penetration of our perceptual experience is still the most plausible account". In: *Consciousness and Cognition* 47, pp. 26–37. DOI: 10.1016/j.concog.2016.09.005.

Nichols, T. E. and A. P. Holmes (2001). "Nonparametric Permutation Tests For Functional Neuroimaging : A Primer with Examples". In: *Human Brain Mapping* 25.15, pp. 1–25. DOI: 10.1002/hbm.1058.

Nili, H. et al. (2014). "A Toolbox for Representational Similarity Analysis". In: *PLoS Computational Biology* 10.4, e1003553. DOI: 10.1371/journal.pcbi.1003553.

Nilsson, R. et al. (2006). "Evaluating feature selection for SVMs in high dimensions". In: *Lecture Notes in Computer Science* 4212, p. 719. DOI: 10.1016/S0377-2217(02)00911-6.

No, R. Lorente de (1947). "A study of nerve physiology". In: *Studies from the Rockefeller institute for medical research. Reprints. Rockefeller Institute for Medical Research* 131, pp. 1–496.

Noble, W. S. (2006). "What is a support vector machine?" In: *Nature Biotechnology* 24.12, pp. 1565–1567. DOI: 10.1038/nbt1206-1565.

Nobre, A. C. (2001). "Orienting attention to instants in time". In: *Neuropsychologia* 39.12, pp. 1317–1328. DOI: 10.1016/S0028-3932(01)00120-8.

Noirhomme, Q. et al. (2014). "Biased binomial assessment of cross-validated estimation of classification accuracies illustrated in diagnosis predictions". In: *NeuroImage: Clinical* 4, pp. 687–694. DOI: 10.1016/j.nicl.2014.04.004.

Nunez, P. L. and R. B. Silberstein (2000). "On the relationship of synaptic activity to macroscopic measurements: Does co-registration of EEG with fMRI make sense?" In: *Brain Topography* 13.2, pp. 79–96. DOI: 10.1023/A:1026683200895.

Öhman, A., D. Lundqvist, and F. Esteves (2001). "The Face in the Crowd Revisited: A Threat Advantage With Schematic Stimuli". In: *Journal of Personality and Social Psychology* 80.3, pp. 381–396. DOI: 10.1037//0022-3514.80.3.381.

Okazaki, Y. O. et al. (2015). "Real-time MEG neurofeedback training of posterior alpha activity modulates subsequent visual detection performance". In: *NeuroImage* 107, pp. 323–332. DOI: 10.1016/j.neuroimage.2014.12.014.

Oliva, A. and A. Torralba (2001). "Modeling the Shape of the Scene : A Holistic Representation of the Spatial Envelope". In: *International Journal of Computer Vision* 42.3, pp. 145–175. DOI: 10.1023/A:1011139631724.

Oliveira, L. et al. (2013). "Emotion and attention interaction: a trade-off between stimuli relevance, motivation and individual differences". In: *Frontiers in Human Neuroscience* 7, p. 364. DOI: 10.3389/FNHUM.2013.00364.

Oostenveld, R. et al. (2011). "FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data". In: *Computational Intelligence and Neuroscience* 2011, p. 156869. DOI: 10.1155/2011/156869.

Padmala, S and L Pessoa (2008). "Affective learning enhances visual detection and responses in primary visual cortex". In: *The Journal of Neuroscience* 28.24, pp. 6202–6210. DOI: 10.1523/JNEUROSCI.1233-08.2008.

Pantazis, D. et al. (2017). "Decoding the orientation of contrast edges from MEG evoked and induced responses". In: *NeuroImage* 180(Pt A), pp. 267–279. DOI: 10.1016/j.neuroimage.2017.07.022.

Peelen, M. V. and P. E. Downing (2017). "Category selectivity in human visual cortex: Beyond visual object recognition". In: *Neuropsychologia* 105, pp. 177–183. DOI: 10.1016/j.neuropsychologia.2017.03.033.

Pegna, A. J. et al. (2005). "Discriminating emotional faces without primary visual cortices involves the right amygdala". In: *Nature Neuroscience* 8.1, pp. 24–25. DOI: 10.1038/nn1364.

Pelli, D. G. (1997). "The VideoToolbox software for visual psychophysics: Transforming numbers into movies". In: *Spatial Vision* 10.4, pp. 437–442. DOI: 10.1163/156856897X00366.

Pereira, F. and M. Botvinick (2011). "Information mapping with pattern classifiers: a comparative study". In: *NeuroImage* 56.2, pp. 476–496. DOI: 10.1016/j.neuroimage.2010.05.026.Information.

Pereira, F., T. Mitchell, and M. Botvinick (2009). "Machine learning classifiers and fMRI : A tutorial overview". In: *NeuroImage* 45.1, S199–S209. DOI: 10.1016/j.neuroimage.2008.11.007.

Perrett, D. I., E. T. Rolls, and W. Caan (1982). "Visual neurones responsive to faces in the monkey temporal cortex". In: *Experimental Brain Research* 47.3, pp. 329–342. DOI: 10.1007/BF00239352.

Perry, G. (2016). "The visual gamma response to faces reflects the presence of sensory evidence and not awareness of the stimulus". In: *Royal Society Open Science* 3.3, p. 150593. DOI: 10.1098/rsos.150593.

Perry, G. and K. D. Singh (2014). "Localizing evoked and induced responses to faces using magnetoencephalography". In: *European Journal of Neuroscience* 39.9, pp. 1517–1527. DOI: 10.1111/ejn.12520.

Pessoa, L et al. (2002a). "Neural processing of emotional faces requires attention." In: *Proceedings of the National Academy of Sciences of the United States of America* 99.17, pp. 11458–11463. DOI: 10.1073/pnas.172403899.

Pessoa, L. (2005). "To what extent are emotional visual stimuli processed without attention and awareness?" In: *Current Opinion in Neurobiology* 15.2, pp. 188–196. DOI: 10.1016/j.conb.2005.03.002.

— (2008). "On the relationship between emotion and cognition". In: *Nature Reviews Neuroscience* 9, pp. 148–158.

— (2010). "Emotion and attention effects: is it all a matter of timing? Not yet". In: *Frontiers in Human Neuroscience* 4.September, pp. 1–5. DOI: 10.3389/fnhum.2010.00172.

Pessoa, L. and R. Adolphs (2010). "Emotion processing and the amygdala: from a 'low road' to 'many roads' of evaluating biological significance". In: *Nature Reviews Neuroscience* 11.11, pp. 773–783. DOI: 10.1038/nrn2920.

— (2011). "Emotion and the brain: Multiple roads are better than one". In: *Nature Reviews Neuroscience* 12.7, p. 425. DOI: 10.1038/nrn2920-c2.

Pessoa, L., S. Kastner, and L. G. Ungerleider (2002b). "Attentional control of the processing of neutral and emotional stimuli". In: *Cognitive Brain Research* 15.1, pp. 31–45. DOI: 10.1016/S0926-6410(02)00214-8.

Pessoa, L., S. Kastner, and L. G. Ungerleider (2003). "Neuroimaging studies of attention: from modulation of sensory processing to top-down control." In: *Journal of Neuroscience* 23.10, pp. 3990–3998. DOI: 23/10/3990[pii].

Pessoa, L., S. Padmala, and T. Morland (2005a). "Fate of unattended fearful faces in the amygdala is determined by both attentional resources and cognitive modulation". In: *NeuroImage* 28.1, pp. 249–255. DOI: 10.1016/J.NEUROIMAGE.2005.05.048.

Pessoa, L., S. Japee, and L. G. Ungerleider (2005b). "Visual Awareness and the Detection of Fearful Faces". In: *Emotion* 5.2, pp. 243–247. DOI: 10.1037/1528-3542.5.2.243.

Pessoa, L., S. Japee, and D. Sturman (2006). "Target Visibility and Visual Awareness Modulate Amygdala Responses to Fearful Faces". In: *Cerebral Cortex* 16.March, pp. 366–375. DOI: 10.1093/cercor/bhi115.

Peyrin, C. et al. (2010). "The neural substrates and timing of top-down processes during coarse-to-fine categorization of visual scenes: a combined fMRI and ERP study." In: *Journal of Cognitive Neuroscience* 22.1994, pp. 2768–2780. DOI: 10.1162/jocn.2010.21424.

Phan, K. L. et al. (2002). "Functional neuroanatomy of emotion: a meta-analysis of emotion activation studies in PET and fMRI." In: *NeuroImage* 16.2, pp. 331–48. DOI: 10.1006/nimg.2002.1087.

Phipson, B. and G. K. Smyth (2010). "Permutation P-values Should Never Be Zero : Calculating Exact P-values When Permutations Are Randomly Drawn". In: *Statistical Applications in Genetics and Molecular Biology* 9.1, p. 39. DOI: 10.2202/1544-6115.1585.

Pichon, S., B. De Gelder, and J. Grèzes (2012). "Threat prompts defensive brain responses independently of attentional control". In: *Cerebral Cortex* 22.2, pp. 274–285. DOI: 10.1093/cercor/bhr060.

Piepers, D. W. and R. A. Robbins (2012). "A review and clarification of the terms "holistic," "configural," and "relational" in the face perception literature". In: *Frontiers in Psychology* 3, pp. 1–11. DOI: 10.3389/fpsyg.2012.00559.

Pitcher, D., V. Walsh, and B. Duchaine (2011). "The role of the occipital face area in the cortical face perception network". In: *Experimental Brain Research* 209.4, pp. 481–493. DOI: 10.1007/s00221-011-2579-1.

Poldrack, R. A. (2011). "Inferring mental states from neuroimaging data: From reverse inference to large-scale decoding". In: *Neuron* 72.5, pp. 692–697. DOI: 10.1016/j.neuron.2011.11.001.

Pourtois, G. et al. (2005). "Enhanced extrastriate visual response to bandpass spatial frequency filtered fearful faces: Time course and topographic evoked-potentials mapping". In: *Human Brain Mapping* 26.1, pp. 65–79. DOI: 10.1002/hbm.20130.

Pourtois, G. et al. (2006). "Neural systems for orienting attention to the location of threat signals: An event-related fMRI study". In: *NeuroImage* 31.2, pp. 920–933. DOI: 10.1016/J.NEUROIMAGE.2005.12.034.

Pourtois, G. et al. (2010). "Temporal precedence of emotion over attention modulations in the lateral amygdala: Intracranial ERP evidence from a patient with temporal lobe epilepsy." In: *Cognitive, Affective & Behavioral Neuroscience* 10.1, pp. 83–93. DOI: 10.3758/CABN.10.1.83.

Pourtois, G., A. Schettino, and P. Vuilleumier (2013). "Brain mechanisms for emotional influences on perception and attention: What is magic and what is not". In: *Biological Psychology* 92.3, pp. 492–512. DOI: 10.1016/j.biopsycho.2012.02.007.

Puls, S. and K. Rothermund (2018). "Attending to emotional expressions: no evidence for automatic capture in the dot-probe task". In: *Cognition and Emotion* 32.3, pp. 450–463. DOI: 10.1080/02699931.2017.1314932.

Pyles, J. A. et al. (2013). "Explicating the Face Perception Network with White Matter Connectivity". In: *PLoS ONE* 8.4, pp. 1–12. DOI: 10.1371/journal.pone.0061611.

Rajaei, K. et al. (2018). "Beyond Core Object Recognition: Recurrent processes account for object recognition under occlusion". In: *bioRxiv* April 2018. DOI: 10.1101/302034.

Rajimehr, R. et al. (2011). "The "Parahippocampal Place Area" Responds Preferentially to High Spatial Frequencies in Humans and Monkeys". In: *PLoS Biology* 9.4, e1000608. DOI: 10.1371/journal.pbio.1000608.

Ramkumar, P. et al. (2016). "Visual information representation and rapid scene categorization are simultaneous across cortex: An MEG study". In: *NeuroImage* 134, pp. 295–304. DOI: 10.1016/j.neuroimage.2016.03.027.

Rawlinson, D. and G. Kowadlo (2017). "Computational Neuroscience Offers Hints for More General Machine Learning". In: *International Conference on Artificial General Intelligence*. Melbourne: Springer, Cham, pp. 123–132. DOI: 10.1007/978-3-319-63703-7{_}12.

Reiss, J. E. and J. E. Hoffman (2007). "Disruption of early face recognition processes by object substitution masking". In: *Visual Cognition* 15.7, pp. 789–798. DOI: 10.1080/13506280701307035.

Rice, G. E. et al. (2014). "Low-Level Image Properties of Visual Objects Predict Patterns of Neural Response across Category-Selective Regions of the Ventral Visual Pathway". In: *Journal of Neuroscience* 34.26, pp. 8837–8844. DOI: 10.1523/JNEUROSCI.5265-13.2014.

Richler, J. J. and I. Gauthier (2014). "A meta-analysis and review of holistic face processing." In: *Psychological Bulletin* 140.5, pp. 1281–1302. DOI: 10.1037/a0037004.

Ritchie, J. B. and T. A. Carlson (2016). "Neural Decoding and "Inner" Psychophysics: A Distance-to-Bound Approach for Linking Mind, Brain, and Behavior". In: *Frontiers in Neuroscience* 10, p. 190. DOI: 10.3389/fnins.2016.00190.

Ritchie, J. B., D. M. Kaplan, and C. Klein (2017). "Decoding the brain: Neural representation and the limits of multivariate pattern analysis in cognitive neuroscience". In: *The British Journal for the Philosophy of Science* axx023. DOI: 10.1101/127233.

Rivolta, D. et al. (2012). "An early category-specific neural response for the perception of both places and faces". In: *Cognitive Neuroscience* 3.1, pp. 45–51. DOI: 10.1080/17588928.2011.604726.

Rivolta, D., A. Puce, and M. A. Williams (2016). "Editorial : Facing the Other : Novel Theories and Methods in Face Perception Research". In: *Frontiers in Human Neuroscience* 10, p. 32. DOI: 10.3389/fnhum.2016.00032.

Riwkes, S., A. Goldstein, and E. Gilboa-Schechtman (2015). "The temporal unfolding of face processing in social anxiety disorder–a MEG study." In: *NeuroImage. Clinical* 7, pp. 678–87. DOI: 10.1016/j.nicl.2014.11.002.

Robinson, S. and J. Vrba (1999). "Functional neuroimaging by synthetic aperture magnetometry (SAM)". In: *Recent Advances in Biomagnetism*. Ed. by T. Yoshimoto et al. Sendai: Tohoku University Press, pp. 302–305. DOI: 10.4236/jbnb.2011.225065.

Rodriguez, V. et al. (2013). "Absence of Face-specific Cortical Activity in the Complete Absence of Awareness: Converging Evidence from Functional Magnetic

Resonance Imaging and Event- related Potentials". In: *Journal of Cognitive Neuroscience* 24.2, pp. 396–415. DOI: 10.1162/jocn.

Roijendijk, L. et al. (2013). "Exploring the Impact of Target Eccentricity and Task Difficulty on Covert Visual Spatial Attention and Its Implications for Brain Computer Interfacing". In: *PLoS ONE* 8.12, e80489. DOI: 10.1371/journal.pone.0080489.

Rossion, B. and S. Caharel (2011). "ERP evidence for the speed of face categorization in the human brain : Disentangling the contribution of low-level visual cues from face perception". In: *Vision Research* 51.12, pp. 1297–1311. DOI: 10.1016/j.visres.2011.04.003.

Rousselet, G. A., O. R. Joubert, and M Fabre-Thorpe (2005). "How long to get to the "gist" of real-world natural scenes?" In: *Visual Cognition* 12.6, pp. 852–877. DOI: 10.1080/13506280444000553.

Rust, N. C. and J. J. DiCarlo (2010). "Selectivity and Tolerance ("Invariance") Both Increase as Visual Information Propagates from Cortical Area V4 to IT". In: *Journal of Neuroscience* 30.39, pp. 12978–12995. DOI: 10.1523/JNEUROSCI.0179-10.2010.

Sacchett, C. and G. W. Humphreys (1992). "Calling a squirrel a squirrel but a canoe a wigwam: a category-specific deficit for artefactual objects and body parts". In: *Cognitive Neuropsychology* 9.1, pp. 73–86. DOI: 10.1080/02643299208252053.

Said, C. P. et al. (2018). "Distributed representations of dynamic facial expressions in the superior temporal sulcus". In: *Journal of Vision* 10.5, pp. 1–12. DOI: 10.1167/10.5.11.Introduction.

Sandberg, K. et al. (2010). "Measuring consciousness: Is one measure better than the other?" In: *Consciousness and Cognition* 19.4, pp. 1069–1078. DOI: 10.1016/j.concog.2009.12.013.

Santesso, D. L. et al. (2008). "Electrophysiological correlates of spatial orienting towards angry faces: A source localization study". In: *Neuropsychologia* 46.5, pp. 1338–1348. DOI: 10.1016/J.NEUROPSYCHOLOGIA.2007.12.013.

Sarvas, J. (1987). "Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem". In: *Physics in Medicine and Biology* 32.1, pp. 11–22. DOI: 10.1088/0031-9155/32/1/004.

Sassi, F. et al. (2014). "Task Difficulty and Response Complexity Modulate Affective Priming by Emotional Facial Expressions". In: *Quarterly Journal of Experimental Psychology* 67.5, pp. 861–871. DOI: 10.1080/17470218.2013.836233.

Sato, M. et al. (2018). "Information spreading by a combination of MEG source estimation and multivariate pattern classification". In: *PLOS ONE* 13.6, e0198806. DOI: 10.1371/journal.pone.0198806.

Sato, N et al. (1999). "Different time course between scene processing and face processing: a MEG study." In: *Neuroreport* 10.17, pp. 3633–7. DOI: 10.1097/00001756-199911260-00031.

Schindler, A. and A. Bartels (2016). "Visual high-level regions respond to high-level stimulus content in the absence of low-level confounds". In: *NeuroImage* 132, pp. 520–525. DOI: 10.1016/j.neuroimage.2016.03.011.

Schlossmacher, I. et al. (2017). "No differential effects to facial expressions under continuous flash suppression: An event-related potentials study". In: *NeuroImage* 163, pp. 276–285. DOI: 10.1016/j.neuroimage.2017.09.034.

Scholte, H. S. (2018). "Fantastic DNimals and where to find them". In: *NeuroImage* 180, pp. 112–113. DOI: 10.1016/j.neuroimage.2017.12.077.

Scholte, H. S., A. W. M. Smeulders, and V. A. F. Lamme (2009). "Brain responses strongly correlate with Weibull image statistics when processing natural images". In: *Journal of Vision* 9.4, pp. 1–15. DOI: 10.1167/9.4.29.Introduction.

Schupp, H. T. et al. (2004). "The Facilitated Processing of Threatening Faces : An ERP Analysis". In: *Emotion* 4.2, pp. 189–200. DOI: 10.1037/1528-3542.4.2.189.

Schyns, P. G. and A. Oliva (1994). "Evidence for Time- and Spatial-Scale-Dependent Scene Recognition". In: *Psychological Science* 5.4, pp. 195–201.

Scott, D. W. (1992). *Multivariate Density Estimation : Theory, Practice, and Visualization*. New York: Wiley.

Seeliger, K et al. (2017). "Convolutional neural network-based encoding and decoding of visual object recognition in space and time". In: *NeuroImage* 180(Pt A), pp. 1–14. DOI: 10.1016/j.neuroimage.2017.07.018.

Seger, C. A. and E. J. Peterson (2013). "Categorization = decision making + generalization". In: *Neuroscience & Biobehavioral Reviews* 37.7, pp. 1187–1200. DOI: 10.1016/j.neubiorev.2013.03.015.

Sekihara, K. et al. (2004). "Asymptotic SNR of Scalar and Vector Minimum-Variance Beamformers for Neuromagnetic Source Reconstruction". In: *IEEE Transactions on Biomedical Engineering* 51.10, pp. 1726–1734. DOI: 10.1109/TBME.2004.827926.

Silva, F. Lopes da (2010). "Electrophysiological Basis of MEG Signals". In: *MEG : an introduction to methods*. Ed. by P. C. Hansen, M. L. Kringelbach, and R. Salmelin. Oxford: Oxford University Press.

Silvert, L. et al. (2007). "Influence of attentional demands on the processing of emotional facial expressions in the amygdala". In: *NeuroImage* 38.2, pp. 357–366. DOI: 10.1016/J.NEUROIMAGE.2007.07.023.

Singh, K. (2006). "Magnetoencephalography". In: *Methods in Mind*. Cambridge: MIT Press, pp. 291–326.

Singh, K. D., G. R. Barnes, and A. Hillebrand (2003). "Group imaging of task-related changes in cortical synchronisation using nonparametric permutation testing". In: *NeuroImage* 19, pp. 1589–1601. DOI: 10.1016/S1053-8119(03)00249-0.

Sinha, P. (2002). "Qualitative Representations for Recognition". In: *BMCV '02 Proceedings of the Second International Workshop on Biologically Motivated Computer Vision*. London: Springer, pp. 249–662. DOI: 10.1167/1.3.298.

Smith, S. M. (2002). "Fast robust automated brain extraction." In: *Human Brain Mapping* 17.3, pp. 143–55. DOI: 10.1002/hbm.10062.

Smith, S. M. and T. E. Nichols (2018). "Statistical Challenges in " Big Data " Human Neuroimaging". In: *Neuron* 97.2, pp. 263–268. DOI: 10.1016/j.neuron.2017.12.018.

Song, C. and H. Yao (2016). "Unconscious processing of invisible visual stimuli". In: *Scientific Reports* 6, p. 38917. DOI: 10.1038/srep38917.

Srinivasan, R., J. D. Golomb, and A. M. Martinez (2016). "A Neural Basis of Facial Action Recognition in Humans". In: *Journal of Neuroscience* 36.16, pp. 4434–4442. DOI: 10.1523/JNEUROSCI.1704-15.2016.

Stefanics, G. et al. (2012). "Processing of unattended facial emotions: A visual mismatch negativity study". In: *NeuroImage* 59.3, pp. 3042–3049. DOI: 10.1016/j.neuroimage.2011.10.041.

Stelzer, J., Y. Chen, and R. Turner (2013). "Statistical inference and multiple testing correction in classification-based multi-voxel pattern analysis (MVPA): Random permutations and cluster size control". In: *NeuroImage* 65, pp. 69–82. DOI: 10.1016/j.neuroimage.2012.09.063.

Sterzer, P., L. Jalkanen, and G. Rees (2009). "Electromagnetic responses to invisible face stimuli during binocular suppression". In: *NeuroImage* 46.3, pp. 803–808. DOI: 10.1016/j.neuroimage.2009.02.046.

Straube, T., M. Mothes-Lasch, and W. H. R. Miltner (2011). "Neural mechanisms of the automatic processing of emotional information from faces and voices". In: *British Journal of Psychology* 102.4, pp. 830–848. DOI: 10.1111/j.2044-8295.2011.02056.x.

Studebaker, G. (1985). "A "rationalized" arcsine transform". In: *Journal of speech and hearing research* 28, pp. 455–462. DOI: 10.1044/jshr.2803.455.

Su, L. et al. (2012). "Spatiotemporal Searchlight Representational Similarity Analysis in EMEG Source Space". In: *Second International Workshop on Pattern Recognition in NeuroImaging Spatiotemporal*. DOI: 10.1109/PRNI.2012.26.

Sussillo, D. (2014). "Neural circuits as computational dynamical systems". In: *Current Opinion in Neurobiology* 25, pp. 156–163. DOI: 10.1016/J.CONB.2014.01.008.

Suzuki, A et al. (2011). "Sustained happiness? Lack of repetition suppression in right-ventral visual cortex for happy faces". In: *Social Cognitive and Affective Neuroscience* 6, pp. 424–441. DOI: 10.1093/scan/nsq058.

Szczepanowski, R. and L. Pessoa (2007). "Fear perception: Can objective and subjective awareness measures be dissociated?" In: *Journal of Vision* 7.4, pp. 1–17. DOI: 10.1167/7.4.10.

Tamietto, M. and B. De Gelder (2010). "Neural bases of the non-conscious perception of emotional signals". In: *Nature Reviews Neuroscience* 11.10, pp. 697–709. DOI: 10.1038/nrn2889.

Tang, H. and G. Kreiman (2017). "Recognition of Occluded Objects". In: *Computational and Cognitive Neuroscience of Vision. Cognitive Science and Technology*. Ed. by Q. Zhao. Singapore: Springer. DOI: 10.1007/978-981-10-0213-7.

Teufel, C. (2018). "Sensory Neuroscience: Linking Dopamine, Expectation, and Hallucinations". In: *Current Biology* 28.4, R142–R143. DOI: 10.1016/j.cub.2018.01.003.

Teufel, C. and B. Nanay (2017). "How to (and how not to) think about top-down influences on visual perception". In: *Consciousness and Cognition* 47, pp. 17–25. DOI: 10.1016/j.concog.2016.05.008.

Teufel, C. et al. (2013). "What is social about social perception research?" In: *Frontiers in Integrative Neuroscience* 6, p. 128. DOI: 10.3389/fnint.2012.00128.

Todd, M. T., L. E. Nystrom, and J. D. Cohen (2013). "Confounds in multivariate pattern analysis : Theory and rule representation case study". In: *NeuroImage* 77, pp. 157–165. DOI: 10.1016/j.neuroimage.2013.03.039.

Tong, F. and M. S. Pratte (2012). "Decoding Patterns of Human Brain Activity". In: *Annual Review ofPsychology is* 63, 483–509. DOI: 10.1146/annurev-psych-120710-100412.

Tonin, L, R Leeb, and J del R Millán (2012). "Time-dependent approach for single trial classification of covert visuospatial attention". In: *Journal of Neural Engineering* 9.4, p. 045011. DOI: 10.1088/1741-2560/9/4/045011.

Tottenham, N. et al. (2009). "The NimStim set of facial expressions: judgments from untrained research participants." In: *Psychiatry research* 168.3, pp. 242–9. DOI: 10.1016/j.psychres.2008.05.006.

Treder, M. S. et al. (2011). *Brain-computer interfacing using modulations of alpha activity induced by covert shifts of attention*. Tech. rep., p. 24. DOI: 10.1186/1743-0003-8-24.

Tsao, D. Y. et al. (2006). "A Cortical Region Consisting Entirely of Face-Selective Cells". In: *Science* 311.February, pp. 670–675.

Tsuchiya, N. et al. (2008). "Decoding Face Information in Time , Frequency and Space from Direct Intracranial Recordings of the Human Brain". In: *PLoS ONE* 3.12, pp. 1–17. DOI: 10.1371/journal.pone.0003892.

Turner, B. M., S. Miletić, and B. U. Forstmann (2018). "Outlook on deep neural networks in computational cognitive neuroscience". In: *NeuroImage* 180, pp. 117–118. DOI: 10.1016/j.neuroimage.2017.12.078.

Tzourio-Mazoyer, N et al. (2002). "Automated Anatomical Labeling of Activations in SPM Using a Macroscopic Anatomical Parcellation of the MNI MRI Single-Subject Brain". In: *NeuroImage* 15, pp. 273–289. DOI: 10.1006/nimg.2001.0978.

Ungerleider, L. G. and J. V. Haxby (1994). "'What' and 'where' in the human brain." In: *Current opinion in neurobiology* 4.2, pp. 157–65. DOI: 10.1016/0959-4388(94)90066-3.

Van Veen, B. et al. (1997). "Localization of brain electrical activity via linearly constrained minimum variance spatial filtering". In: *IEEE Transactions on Biomedical engineering* 44.9, pp. 867–880. DOI: 10.1109/10.623056.

VanRullen, R. (2017). "Perception science in the age of deep neural networks". In: *Frontiers in Psychology* 8, p. 142. DOI: 10.3389/fpsyg.2017.00142.

VanRullen, R. and S. J. Thorpe (2001). "The Time Course of Visual Processing: From Early Perception to Decision-Making". In: *Journal of Cognitive Neuroscience* 13.4, pp. 454–461. DOI: 10.1162/08989290152001880.

Varoquaux, G. and B. Thirion (2014). "How machine learning is shaping cognitive neuroimaging". In: *GigaScience* 3.28, pp. 1–7.

Vaziri-Pashkam, M. and Y. Xu (2017). "Goal-Directed Visual Processing Differentially Impacts Human Ventral and Dorsal Visual Representations". In: *The Journal of Neuroscience* 37.36, pp. 8767–8782. DOI: 10.1523/JNEUROSCI.3392-16.2017.

Vida, M. D. et al. (2017). "Spatiotemporal dynamics of similarity-based neural representations of facial identity". In: *Proceedings of the National Academy of Sciences* 114.2, pp. 388–393. DOI: 10.6084/m9.figshare.4233107.v1.

Vidaurre, D. et al. (2018). "Temporally unconstrained decoding reveals consistent but time-varying stages of stimulus processing". In: *bioRxiv*, p. 260943. DOI: 10.1101/260943.

Visconti Di Oleggio Castello, M. et al. (2017). "Familiarity facilitates feature-based face processing". In: *PLoS One* 12.6, e0178895. DOI: 10.1371/journal.pone.0178895.

Vrba, J. (2002). "Magnetoencephalography: The art of finding a needle in a haystack". In: *Physica C: Superconductivity and its Applications* 368.1-4, pp. 1–9. DOI: 10.1016/S0921-4534(01)01131-5.

Vrba, J and S. E. Robinson (2001). "Signal processing in magnetoencephalography". In: *Methods* 25.2, pp. 249–271. DOI: 10.1006/meth.2001.1238.

Vuilleumier, P. (2005). "How brains beware: Neural mechanisms of emotional attention". In: *Trends in Cognitive Sciences* 9.12, pp. 585–594. DOI: 10.1016/j.tics.2005.10.011.

Vuilleumier, P. and R. Righart (2012). *Attention and Automaticity in Processing Facial Expressions*. September 2018, pp. 1–42. DOI: 10.1093/oxfordhb/9780199559053.013.0023.

Vuilleumier, P. et al. (2001). "Effects of Attention and Emotion on Face Processing in the Human Brain: An Event-Related fMRI Study". In: *Neuron* 30.3, pp. 829–841. DOI: 10.1016/S0896-6273(01)00328-2.

Vuilleumier, P. et al. (2003). "Distinct spatial frequency sensitivities for processing faces and emotional expressions." In: *Nature neuroscience* 6.6, pp. 624–631. DOI: 10.1038/nn1057.

Waal, F. B. de (2000). "Primates–a natural heritage of conflict resolution." In: *Science* 289, pp. 586–90. DOI: 10.1126/science.289.5479.586.

Walther, A. et al. (2016). "Reliability of dissimilarity measures for multi-voxel pattern analysis". In: *NeuroImage* 137, pp. 188–200. DOI: 10.1016/j.neuroimage.2015.12.012.

Walther, D. B. et al. (2009). "Natural Scene Categories Revealed in Distributed Patterns of Activity in the Human Brain". In: *Journal of Neuroscience* 29.34, pp. 10573–10581. DOI: 10.1523/JNEUROSCI.0559-09.2009.

Walther, D. B. et al. (2011). "Simple line drawings suffice for functional MRI decoding of natural scene categories." In: *Proceedings of the National Academy of Sciences of the United States of America* 108.23, pp. 9661–9666. DOI: 10.1073/pnas.1015666108.

Wang, X. et al. (2016). "The Hierarchical Structure of the Face Network Revealed by Its Functional Connectivity Pattern". In: *Journal of Neuroscience* 36.3, pp. 890–900. DOI: 10.1523/JNEUROSCI.2789-15.2016.

Wardle, S. G. et al. (2016). "Perceptual similarity of visual patterns predicts dynamic neural activation patterns measured with MEG". In: *NeuroImage* 132, pp. 59–70. DOI: 10.1016/j.neuroimage.2016.02.019.

Warrington, E. K. and T Shallice (1984). "Category specific semantic impairments." In: *Brain : a journal of neurology* 107 (Pt 3), pp. 829–54.

Watson, D. M., T. Hartley, and T. J. Andrews (2014). "Patterns of response to visual scenes are linked to the low-level properties of the image." In: *NeuroImage* 99, pp. 402–10. DOI: 10.1016/j.neuroimage.2014.05.045.

Watson, D. M. et al. (2016). "Patterns of neural response in scene-selective regions of the human brain are affected by low-level manipulations of spatial frequency". In: *NeuroImage* 124, pp. 107–117. DOI: 10.1016/j.neuroimage.2015.08.058.

Watson, D. M., T. J. Andrews, and T. Hartley (2017). "A data driven approach to understanding the organization of high-level visual cortex". In: *Scientific Reports* 7.1, p. 3596. DOI: 10.1038/s41598-017-03974-5.

Wegrzyn, M. et al. (2015). "Investigating the brain basis of facial expression perception using multi-voxel pattern analysis". In: *Cortex* 69, pp. 131–140. DOI: 10.1016/j.cortex.2015.05.003.

Wegrzyn, M. et al. (2017). "Mapping the emotional face . How individual face parts contribute to successful emotion recognition". In: *PLoS ONE* 12.5, pp. 1–15.

Weymar, M. et al. (2011). "The face is more than its parts - Brain dynamics of enhanced spatial attention to schematic threat". In: *NeuroImage* 58.3, pp. 946–954. DOI: 10.1016/j.neuroimage.2011.06.061.

Wierzchoń, M. et al. (2014). "Different subjective awareness measures demonstrate the influence of visual identification on perceptual awareness ratings". In: *Consciousness and Cognition* 27.1, pp. 109–120. DOI: 10.1016/j.concog.2014.04.009.

Willenbockel, V. et al. (2010). "Controlling low-level image properties: The SHINE toolbox". In: *Behavior Research Methods* 42.3, pp. 671–684. DOI: 10.3758/BRM.42.3.671.

Williams, L. M. et al. (2004). "Mapping the Time Course of Nonconscious and Conscious Perception of Fear: An Integration of Central and Peripheral Measures". In: *Human Brain Mapping* 21.2, pp. 64–74. DOI: 10.1002/hbm.10154.

Williams, M. A., S. Dang, and N. G. Kanwisher (2007). "Only some spatial patterns of fMRI response are read out in task performance". In: *Nature Neuroscience* 10.6, pp. 685–686. DOI: 10.1038/nn1900.

Williams, N. and R. N. Henson (2018). "Recent advances in functional neuroimaging analysis for cognitive neuroscience". In: *Brain and Neuroscience Advances* 2, pp. 1–4. DOI: 10.1177/2398212817752727.

Woodhead, Z. V. J. et al. (2011). "Dissociation of sensitivity to spatial frequency in word and face preferential areas of the fusiform gyrus". In: *Cerebral Cortex* 21.10, pp. 2307–2312. DOI: 10.1093/cercor/bhr008.

Wyatte, D., D. J. Jilk, and R. C. O'Reilly (2014). "Early recurrent feedback facilitates visual object recognition under challenging conditions". In: *Frontiers in Psychology* 5, p. 674. DOI: 10.3389/fpsyg.2014.00674.

Xiao, J. et al. (2010). "SUN Database : Large-Scale Scene Recognition from Abbey to Zoo". In: *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pp. 3485 –3492.

Yamins, D. L. K. and J. J. DiCarlo (2016). "Using goal-driven deep learning models to understand sensory cortex". In: *Nature Neuroscience* 19.3, pp. 356–365. DOI: 10.1038/nn.4244.

Yamins, D. L. K. et al. (2014). "Performance-optimized hierarchical models predict neural responses in higher visual cortex". In: *Proceedings of the National Academy of Sciences* 111.23, 8619–8624. DOI: 10.1073/pnas.1403112111.

Yin, R. K. (1969). "Looking at upside-down faces." In: *Journal of Experimental Psychology* 81.1, pp. 141–145. DOI: 10.1037/h0027474.

Zachariou, V., Z. N. Safiullah, and L. G. Ungerleider (2018). "The Fusiform and Occipital Face Areas Can Process a Nonface Category Equivalently to Faces". In: *Journal of Cognitive Neuroscience* 30.10, pp. 1499–1516. DOI: 10.1162/jocn{_}a{_}01288.

Zhang, H. et al. (2016a). "Face-selective regions differ in their ability to classify facial expressions". In: *NeuroImage* 130, pp. 77–90. DOI: 10.1016/j.neuroimage.2016.01.045.

Zhang, J. et al. (2016b). "Decoding Brain States Based on Magnetoencephalography From Prespecified Cortical Regions." In: *IEEE transactions on bio-medical engineering* 63.1, pp. 30–42. DOI: 10.1109/TBME.2015.2439216.

Zhou, B. et al. (2014). "Learning Deep Features for Scene Recognition using Places Database". In: *Advances in Neural Information Processing Systems 27*, pp. 487–495. DOI: 10.1162/153244303322533223.

Zimmerman, J. E., P. Thiene, and J. T. Harding (1970). "Design and Operation of Stable rf-Biased Superconducting Point-Contact Quantum Devices, and a Note on the Properties of Perfectly Clean Metal Contacts". In: *Journal of Applied Physics* 41.4, pp. 1572–1580. DOI: 10.1063/1.1659074.

Zion-Golumbic, E. and S. Bentin (2007). "Dissociated Neural Mechanisms for Face Detection and Configural Encoding: Evidence from N170 and Induced Gamma-Band Oscillation Effects". In: *Cerebral Cortex* 17.8, pp. 1741–1749. DOI: 10.1093/cercor/bhl100.