Spatiotemporal responses to natural images and their phase-shuffled version in the primary visual cortex

Sepide Movaghati

Department of Biomedical Engineering
McGill University

Montreal, Canada

April 2015

A thesis submitted to McGill University in partial fulfillment of the requirements

for the degree of Master of Biomedical Engineering

ABSTRACT

According to classical models of visual information processing, the visual cortex acts as a hierarchical structure in which information flow is unidirectional and transmitted from the primary visual cortex (V1) to higher visual areas. This model considers V1 as an inflexible processing unit, which filters local basic features of visual input, such as the spatial frequency and orientation of the changes in luminance. Contrary to this classical model, recent studies have reported variant activity in V1 based on the presence of higher-order features within visual inputs. In other words, specific characteristics, such as phase content or spatial coherence, potentially modulate the activity of V1. However, the explicit characteristics and the precise timing of their modulatory effect remain unclear.

In this study, MagnetoEncephaloGraphy (MEG) was used to test whether the spatiotemporal response to natural images of animals is different from the response to their phase-shuffled counterparts. The frequency content and the total contrast of the images were maintained throughout the phase-scrambling approach. Meanwhile, higher-order image statistics, such as Kurtosis, were manipulated. We investigated the effect of this manipulation on the evoked response of V1 by taking advantage of (1) the high temporal resolution of the MEG technique and (2) the Minimum norm source localization method.

The results of this thesis show that the modulation of activity associated with the presentation of natural images appears as early as 50ms post-stimulation. This effect can be observed in the power of the signals recorded by MEG sensors as well as in sources localized in the primary visual cortex. The phase scrambling of visual input increased the amplitude of the early response (30-70ms post-stimulation) and the M170 component (140-200ms post-stimulation). The weaker population response to natural images during the early response is consistent with a model proposing sparse neuronal coding for natural images. The increased M170 activation of V1 to non-natural phase-scrambled images is consistent with the predictive coding model. According to this model, the prediction error in lower order sensory areas increases with decreasing predictability of the sensory input. An increase in the power of gamma frequency (60-80Hz) was observed in response to phase-scrambled images. These findings are in accordance with a previous study that found prediction errors in auditory cortex to be predominantly conveyed via the high (gamma) frequency range. In addition, spatially coherent features in visual input induce a stronger surround suppression effect and reduce the population activity in V1. Therefore, the mechanisms underlying the late modulatory effect on the M170 component may also involve surround suppression in the visual cortex.

Our findings suggest that feedback modulates the neuronal response of the primary visual cortex. Further, they identify V1 as a dynamic filter that can be modulated by the higher-order characteristics of visual input, as opposed to previous models that deem V1 to be an inflexible low-level processing unit.

ABRÉGÉ

Dans le concept traditionnel du traitement de l'information visuelle, le cortex visuel agi comme une structure entièrement hiérarchique dans laquelle l'information se dirige dans une seule direction : du cortex visuel primaire (V1) aux régions visuelles supérieures. Dans ce modèle hiérarchique, V1 est considéré comme une unité de traitement d'information rigide qui sert à filtrer les qualités de base de l'information visuelle comme la fréquence spatiale et la direction des changements de luminance. Cependant, des études récentes ont noté que l'activité dans V1 peut être modulée en réponse aux changements dans la structure visuelle de plus haute niveau, comme la phase ou la cohérence spatiale. Par contre, les attributs et les caractéristiques temporelles de ces effets modulatoires sont présentement inconnues.

Dans notre étude, nous avons utilisé la MagnétoEncéphaloGraphie (MEG) dans le but de tester si la réponse spatiotemporelle à des images naturelles d'animaux est différente de la réponse aux mêmes images dont la phase à été aléatoirement modifiée (brassage de phase). Le contenu en fréquence et le contraste total des images n'est pas affecté par le brassage de phase. Donc, seulement les statistiques d'ordre élevé comme le coefficient de dissymétrie ou le kurtosis ont été manipulées. Nous avons étudié l'effet de cette manipulation sur la réponse évoquée de V1 en prenant avantage de la haute résolution temporelle de la MEG ainsi que la méthode de localisation de source norme minimale.

Nos résultats montrent que la modulation de l'activité dû à la présentation des images naturelles survient au plus tôt 50ms après la stimulation visuelle. Cet effet peut être observée dans la puissance du signale dans le champ magnétique ainsi que dans les sources localisés dans le cortex visuel primaire. Le brassage de phase a augmenté l'amplitude de la réponse rapide (30-70ms après présentation du stimulus) et de la composante M170 (140-200ms après le stimulus). Par contre, nous avons observé une diminution de la réponse rapide suite à la présentation d'images naturelles. Ceci peut être expliqué avec un modèle de code neuronal parsemé. De plus, 'de l'amplification de la composante M170 dans V1 suivant la présentation d'images non-naturelles (phases brassés) est en ligne avec le modèle d'un code neuronal prédictif. Selon ce modèle, quand la prévisibilité de l'entrée visuelle diminue, l'erreur de la prédiction dans les régions sensorielles d'ordre basse augmente, ce qui entraîne une augmentation de la réponse. Nous avons aussi obsérvé une augmentation dans la puissance dans la bande de fréquence gamma (60-80Hz) en réponse aux images brassées en phase, ce qui concorde avec une étude précédente qui avait trouvé que les erreurs de prédiction dans le cortex auditif étaient surtout représentées dans les hautes fréquences (gamma). Finalement, la cohérence spatiale de l'entrée visuelle produit une suppression des alentours et rédut l'activité des cellules dans V1. Donc, l'effet modulatoire sur le composante M170 peut aussi être relié à la suppression des alentours dans le cortex visuel.

Nos résultats suggèrent que la rétroaction dans le traitement de l'information visuel dans le cerveau module la réponse dans le cortex visuel primaire. Ils identifient aussi cette région comme un filtre dynamique qui peut être modulé par des statistiques d'ordre élevéde l'entrée, et non une unité de traitement rigide.

ACKNOWLEDGMENTS

I would like to thank my supervisor, Dr. Amir Shmuel, for his dedicated support and trust in me. I would also thank Dr. Sylvain Baillet for his valuable guidance in the MEG data acquisition and analysis. I was fortunate to work in a place close to a group of Brainstorm developers, Francois Tadel and Elizabeth Anne Bock, who had offered me their generous help at any time I needed it. I am very grateful to Dr. Curtis Baker for providing me with the Psychophysics experiment setup in his lab for my pilot Psychophysics study. I would also thank Dr. Fred Kingdom, for his worthwhile advice on stimulus design, and Dr. Christophe Grova for his comments on MEG data analysis and source localization.

I owe sincere gratitude to Shahab Bakhtiari, Soheila Samiee, Benjamin Morillon, and Pascal Kropf for their kind help during research and thesis completion. I want to thank Debra Dawson for helping me in writing the French abstract and Michelle Rosenstein for her proofreading and editing which greatly helped the readability of my work. Finally, I would like to thank Biomedical Engineering Department, specially Dr. Robert Kearney and Dr. Louis Collins, for their strong support and help during the course of my study and research.

CONTRIBUTION OF AUTHORS:

We intend to submit a shorter version of thesis for a publication in a journal, with the following author contributions:

Authors: Sepide Movaghati, Shahab Bakhtiari, Sylvain Baillet, Amir Shmuel

Study design: Sepide Movaghati, Shahab Bakhtiari, Amir Shmuel

Data acquisition: Sepide Movaghati

Data analysis: Sepide Movaghati, Sylvain Baillet, Amir Shmuel

Writing the manuscript: Sepide Movaghati, Amir Shmuel

Content

List of Figures

List of Acronyms

fMRI	functional Magnetic Resonance Imaging
MEG	Magnetoencephalography
EEG	Electroencephalography
VEP	visual evoked potential
ERP	evoked response potential
BOLD	Blood Oxygen Level Dependent
LO	lateral occipital area
MT	middle temporal area
SNR	signal-to-noise ratio
RMS	root-mean-square
FWHM	full width-half max
ROI	Region of Interest

1. Introduction

1.1. The visual system as a hierarchical structure

The visual cortex is located in the occipital lobe in the posterior region of the brain and is responsible for processing visual information. The sensory input to the visual system is a rich source of information which contains substantial redundancies. Therefore, the first role of the visual system is to identify and extract useful information among this collection of complex input.

A general visual input such as a natural scene, contains several basic low-level features (Graham, 1980; Olshausen and Field, 1996). These features, i.e. the variation in intensity at different spatial frequencies and orientations, are then combined to make structures such as edges, lines, and contours. Further, these different spatial structures merge to form complex objects (Hansen and Hess, 2007). Multiple stages of processing are required to effectively deal with these complex visual inputs. Each processing stage gets progressively more sophisticated and abstract in terms of its representation of the natural visual input (Lennie, 1998; Olshausen, 2013; Tanaka, 1993).

The visual cortex is formed as a hierarchy of areas. Each area has its own specialization in processing visual information (Fig. 1-1). Visual input from the retina is transmitted to the brain. Within the brain, the region responsible for the initial processing of visual information is termed the primary visual cortex (V1). V1 neurons are sensitive to local and basic features of the visual

scene, such as position and orientation, whereas upstream areas are capable of processing more global and complex visual features.

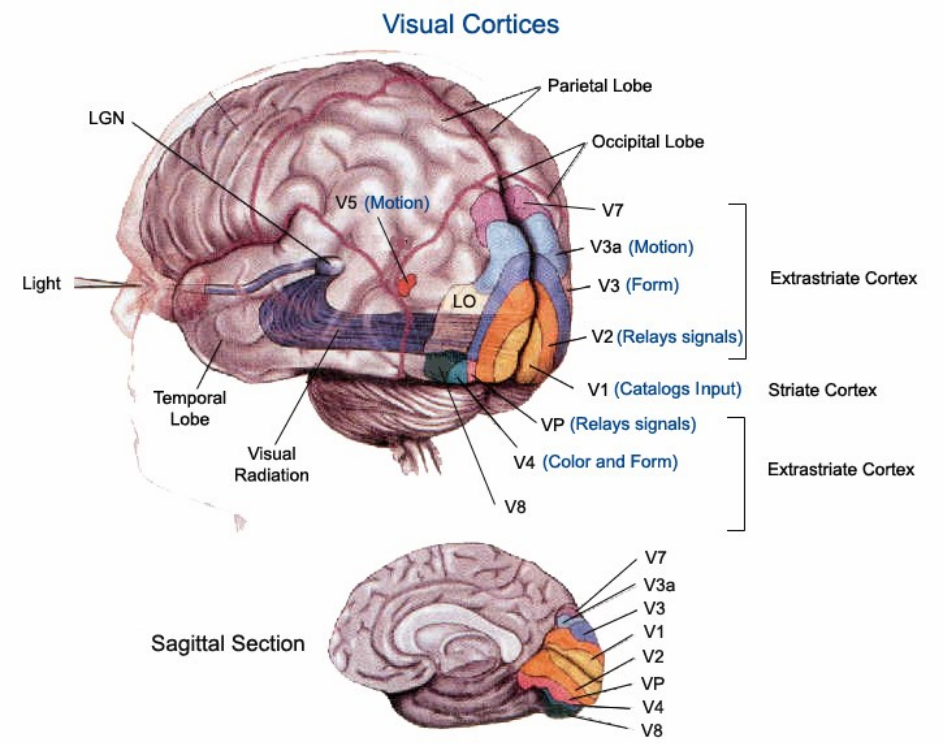

Fig. 1-1: Different visual areas in human brain (Ref: www.techcyn.com/feature.php?id=f2&issue=1).

In general, two parallel processing pathways exist in the visual cortex: the dorsal pathway and the ventral pathway. The dorsal pathway, also known as the “where” pathway, initiates in V1 and continues to the posterior parietal cortex. The dorsal stream is comprised of cortical areas that are specialized for motion and depth perception, such as area MT. The ventral pathway, called the "what" pathway, also commences in V1 but extends toward the inferior temporal cortex (Mishkin et al., 1983). The ventral stream is mainly responsible for processing the form and shape of visual inputs. V4, the lateral occipital area (LO), and the fusiform gyrus are components of the ventral visual pathway and contribute to its functioning (Grill-Spector et al., 1998; Larsson

and Heeger, 2006; Malach et al., 1995; Murray et al., 2002; Sayres and Grill-Spector, 2008; Schwarzlose et al., 2008). V4 is sensitive to coherent shapes and LO shows selectivity to objects (Rainer et al., 2002). Meanwhile, the fusiform face area – along with other areas in the occipito-temporal lobe responsible for object perception and recognition – is a region activated by face stimuli. Higher-order areas in both the dorsal and ventral pathways are sensitive to more global and complex characteristics of a visual scene, such as the overall direction of motion or the appearance of faces and natural objects, respectively.

Based on studies on anatomical connections, response latencies, and neuronal feature selectivity, the flow of information along the visual processing areas is suggested to start from V1, continue to V2, then to V3, V4, LO, and MT up to higher-level inferior-temporal areas (Tjan et al., 2006). Several studies have been done on the characteristics of the responses from this cascade of areas, either in terms of timing or sensitivity (Chen et al., 2010; Freeman and Simoncelli, 2011; Grill-Spector et al., 1998; Hansen and Hess, 2007; Hemond et al., 2007; Henriksson et al., 2009; Rainer et al., 2001, 2002; Tjan et al., 2006) (Chao et al., 1999; Schwarzlose et al., 2008; Tanaka, 1993). As we go higher through these processing stages, we see an increase in feature invariance, closely relating to the visual system's intrinsic uncertainty (Tjan et al., 2006). The notion of feature invariance refers to a processing unit which is invariant to some features in the input. For instance, a system which is invariant to orientation, will respond equally to an input regardless of its orientation. Tjan et al. used this relationship to investigate the change in feature invariance by measuring the response uncertainty using natural images with different levels of signal to noise ratio (SNR). They reported that, keeping the contrast constant, the effect of stimulus SNR on

fMRI responses (%ΔBOLD with respect to ΔSNR) increases as we go far from the retina (Tjan et al., 2006).

In Freeman and Simoncelli (2011), the authors developed another approach to investigate the increasing feature invariance throughout the visual stream. Their model relies on the activity of neuronal ensembles within ventral stream processing. Along the ventral pathway, neuronal receptive fields and eccentricity are proposed to be increasing. Further, by using a novel metameric stimulus in collaboration with the evaluation of receptive field size, this behavioural study successfully identifies visual areas involved in feature representation (Freeman and Simoncelli, 2011).

Although the visual system is formed as a hierarchy of specialized areas, the precise overlap and communication between these areas is still being investigated. In recent decades, it has become clear that the majority of cortical computational work is not carried out by one area at a time. In fact, there exists a large set of reciprocal inter-areal anatomical connections, transferring visual information bi-directionally along feed-forward and feedback streams (Mumford, 1992).

1.2. Feedback and its modulatory role

Based on the classical view of sensory input processing in the brain, the information propagates in a largely sequential, feed-forward manner throughout different cortical areas (Fukushima, 1980; Riesenhuber and Poggio, 1999; Wallis and Rolls, 1997). Recently, it has been shown that this hierarchical feed-forward model is both inadequate for explaining perceptual mechanisms and highly inefficient in terms of information processing. The currently accepted model

recognizes that the computational work mainly results from bidirectional communication both within and between distinct cortical areas (Olshausen, 2013). In vision studies specifically, the existence of anatomical feedback connections serves as foundation for the assumption that top-down modulations in visual processing are functionally significant (Bastos et al., 2015; Ranganath et al., 2004; Tomita et al., 1999).

In Camprodon et al. (2010), the authors study the necessity for feedback projections in a visual recognition task by using Transcranial Magnetic Stimulation (TMS). Human subjects were asked to perform an image categorization task while TMS pulses simultaneously disrupted their V1 neuronal activity at various time points. The researchers report a significant decrease in performance when these TMS pulses were specifically applied at 100 ms and 220 ms post-stimulation. They attribute the first drop in performance, occurring at 100 ms, to interrupted feed-forward transmission of information. Further, the second drop in performance, occurring at 220 ms, is postulated to be due to impaired feedback projects to area V1 (Camprodon et al., 2010). These notions are reinforced by another study, in which the later component of evoked-related potential in EEG, peaking at 200ms post-stimulation, is linked to top-down connections (Garrido et al., 2007).

A recent study uses Dynamic Causal Modeling (DCM) to model directional connectivity between lower and higher order cortical areas in humans during word recognition tasks. The authors observe significant feedback connections from the inferior frontal gyrus to the ventral occipito-temporal and occipital cortex after 100ms post word presentation (Woodhead et al., 2014). In other studies, the authors propose a mechanism for neuronal feedback modulation within the visual system. Based on their proposed model, low spatial frequencies propagate

rapidly to higher frontal areas which then provide a top-down modulatory effect, restricting recognition based on the low spatial frequencies of the input (Bar, 2003; Bar et al., 2006). The presence of modulatory effects within the neuronal networks of the brain is also supported by other studies that focus on attention and memory (Chaumon et al., 2008; Di Russo et al., 2012; Kelly et al., 2008; Poghosyan and Ioannides, 2008).

Considering the growing number of studies reporting this top-down effect, its existence is continuously less questioned. However, the nature and functionality of these modulations remains unclear, and thus, is actively studied. Several models have been proposed to explain the neural mechanism of sensory input processing and perception. Several of these models incorporate top-down effect and in doing so, have successfully accounted for the majority of collected data. In the following section, I present several representative models. Further, I describe studies that aimed at validating these models via data analysis and evidence-based science.

1.3. Neural processing models

1.3.1. Efficient Coding

The neuronal system has adapted to the statistical properties of the sensory input within human environment during a developmental and evolutionary process (Simoncelli and Olshausen, 2001). Therefore, specific properties of the natural input are crucial in determining the neural coding process. One of these properties is the substantial redundancy and correlation within the spatial and temporal domains of sensory input. According to the Efficient Coding Theory

(Attneave, 1954; Barlow, 1961), the neural system aims to both exploit redundancy and efficiently code the useful information presented within the natural input. For instance, when considering visual input, the de-correlation process – termed 'whitening' – is similar to what takes place at the level of retina and Lateral Geniculate Nucleus (LGN). This process aims to reduce the correlation in the input and get rid of redundant or less informative information. However, whitening alone is insufficient for achieving efficient coding (Simoncelli and Olshausen, 2001).

Based on this model, higher processing units are capable of amplifying or attenuating the activity of lower units. This top-down modulation results in a sharper response profile to sensory input (Dumoulin and Hess, 2006). With respect to the visual system specifically, the Efficient Coding Theory can successfully explain neuronal computation from the eye to the first stage of visual processing in V1. However, although this model clearly emphasizes the necessity for efficiently coded information, it fails to elaborate what specific information needs to be coded.

1.3.2. Sparse Coding

In accordance with the notion of evolutionary adaptation of our sensory system, Olshausen and Field propose a model for the representation of natural images using the linear combination of a set of basis functions. In order to find these basis functions, they imposed the constraint of minimizing the number of non-zero coefficients. In other world, they tried to achieve a representation of the natural images with maximum sparseness in the combination of the functions output. The properties of the resulting basis functions resemble the spatially localized Gabor filters consisting of various spatial frequencies and orientations, which matches the receptive fields of simple cells in V1 (Olshausen and Field, 1996, 1997).

Based on this model, sparseness is considered to be a functionally beneficial property of neural coding. Several studies have investigated the relevance of this theory by comparing these theoretical concepts to experimentally attained quantitative data, such as recorded measurements of neuronal activity. In most of the cases, higher sparseness is observed when neural populations respond to natural stimulation rather than noise stimulation (Dumoulin et al., 2008; Froudarakis et al., 2014; Rainer et al., 2001).

Both life-time sparseness and population sparseness can be defined and assessed in physiological recording. Life-time sparseness indicates the amount of response selectivity of a single neuron during a natural vision. Population sparseness corresponds to the number of active neurons in that population of neurons in response to a single stimulation. The observance of life-time and population sparseness within neuronal activity is discussed in a recent study, in which the authors record up to 500 neurons from the mouse V1 area. They report higher population and life-time sparseness in response to natural movies compared to phase-randomized control versions (Froudarakis et al., 2014).

1.3.3. Bayesian Inference Model

As we discussed earlier, 'efficient coding' was successful in modeling neural processing in the retina, LGN, and V1. Similarly, the receptive fields of simple neurons in V1 were successfully modeled by the 'sparse coding' principle. However, a more complex model is required to (1) describe processing beyond V1 and (2) to explain the relationship between different stages of processing. Hierarchical inference models, which are based on Bayes' rule, may be successful in accomplishing these aforementioned goals (Lee and Mumford, 2003).

As previously elaborated, the neural system is not a passive observer. The construction of an inference of the real world using a purely feed-forward system is extremely unlikely, if not impossible. According to the hierarchical inference, an accurate perception of the world requires the combination of sensory data and prior information. In other words, sensory input must be modulated by higher-order cortical structures, in which previous knowledge has been processed and stored, in order to accurately portray ones surroundings.

These inference models propose a clear role for feedback connections: the disambiguation of sensory inputs. Based on Bayesian inference models, the early representation (i.e. the likelihood) of the sensory input aggregates in a feed-forward manner. Subsequently, the higher-order areas disambiguate these weak cues by sending feedback signals (i.e. the priors) to the lower areas. These theories have a strong mathematical framework that can describe the processing in the presence of noisy or incomplete input data (Olshausen, 2013). Further, using fMRI, Naselaris et al. (2009) construct a Bayesian decoder that obtains the signals from the early and anterior visual areas as input and identifies the spatial structure and semantic category of the viewed natural image. This study demonstrates the significant role of prior information in visual processing and image reconstruction (Naselaris et al., 2009). Similarly, another study employed a similar Bayesian decoder in order to model the brain activity while viewing natural movies (Nishimoto et al., 2011).

1.3.4. Predictive Coding

Predictive coding is also a theory based on Bayes' rule (Rao et al., 2007). The main difference between predictive coding and the general Bayesian inference model is the designated role of the feedback signal. Based on Bayesian inference model, feedback connections disambiguate the

representation in lower cortical areas and are dictated by higher-order priors. Therefore, it will facilitate the activation of the neurons which are consistent with higher order representation but suppress the inconsistent activities. While the inference models propose both excitatory and inhibitory actions of feedback modulation, the predictive coding theory suggests that these signals are mainly inhibitory (Olshausen, 2013). Based on the latter model, higher cortical areas transmit their predictions about sensory input in order to cancel out portions of incoming signal that can be explained by higher-level representations. In turn, the lower cortical areas relay information about features of the visual input that could not be predicted by higher-order structures (i.e. the errors) (Mumford, 1992; Rao and Ballard, 1999). In other words, the predictions are conveyed backward through top-down connections and the errors are propagated forward in a bottom-up manner.

The majority of studies that have attempted to validate the predictive coding theory report an association between decreased lower-order cortical activity and increased predictability of the stimulus by higher-order areas (Alink et al., 2010; Chen et al., 2010; Egner et al., 2010; Kok et al., 2012a; Murray et al., 2002; Todorovic and de Lange, 2012). The most important issue in such studies is that detecting the effect of feedback is not straightforward. No clear feature or precise timing is associated with the feedback signal. Moreover, since the top-down effect is not stimulus-locked to the same extent as the bottom-up effect, it is difficult to average results over trials. Unfortunately, the averaging of trials is the most common method to increase the SNR and unmask the effect of interest.

The results of several neuroimaging studies are consistent with the predictive coding theory. One study reports a decrease in the BOLD signal in V1 when the direction of motion of a visual

stimulus is predictable and distinguishable from surrounding illusory motion (Alink et al., 2010). In another fMRI study, the authors employ the predictive coding theory to define the population activity in each cortical area as the sum of activity related to the representational unit (prediction) and the error unit (prediction error). They use two types of stimuli, faces and houses, and measure the BOLD signal of the fusiform face area under different face expectation conditions. They demonstrate that the population activity is best explained by considering prediction and error units in each area as dictated by the predictive coding model (Egner et al., 2010).

Some studies examine the relevance of predictive coding when investigating the process of learning in the brain (den Ouden et al., 2010; den Ouden et al., 2009; Kok et al., 2012b). In Kok et al. (2012a), the authors employ a visual discrimination task preceded by an auditory cue in order to assess and compare responses to expected and unexpected sensory stimuli. They observe a reduction in the BOLD signal in area V1 when the orientation of the grating was expected based on the auditory cue (Kok et al., 2012a). In another study, they report that attention reverses this attenuated effect of the predicted signal (Kok et al., 2012b).

Recent EEG/MEG studies on auditory cortex similarly try to evaluate the role of predictive-coding by comparing predictable versus unpredictable tones and speech (Arnal et al., 2011; Todorovic and de Lange, 2012; Wacongne et al., 2012; Wacongne et al., 2011). One of these studies uses a tone mismatch and omission paradigm to assess the validity of the predictive coding model by comparing the time course of the neuronal response under different conditions. The authors rule out the possibility of activity suppression as a result of adaptation to repetitive tones, stating that predictive coding can better explain the response to the mismatched and omitted tone (Wacongne et al., 2011). Another study reported an increase in the activation of

lower-order area when the written word and the auditory speech are mismatched. The authors of this study describe this modulatory effect as a top-down feedback effect, since the differential response to matched and mismatched stimuli in higher-order areas is observed earlier than the modulatory effect in lower-order areas (Sohoglu et al., 2012).

Another MEG study on the auditory cortex reports an increase in the evoked filed potential after unexpected tone (Todorovic and de Lange, 2012). In addition, while removing the expected response from the unexpected response within 0-350ms post-stimulation, the authors also observed a significant spectrotemporal cluster in 5-9 Hz as well as in gamma band (Todorovic and de Lange, 2012). These results are reinforced by another study, which showed a low-beta (14–15 Hz) / high-gamma (60–80 Hz) coupling regime locally in a multisensory area (area STS) during unexpected conditions (Arnal et al., 2011). The authors of the latter study also observed a slow delta activity (3–4 Hz) in higher areas when the input was correctly predicted. These findings are evidence in support of distinct frequency channels for communicating prediction and prediction errors. Further, it has been hypothesized that top-down predictions are mediated by slower frequencies (alpha/beta) whereas bottom-up prediction errors are conveyed via predominantly high (gamma) frequency ranges (Arnal et al., 2011; Bastos et al., 2015; Bauer et al., 2014; van Kerkoerle et al., 2014; Wang, 2010).

1.4. Natural stimuli

Along with the evolutionary- and developmental-based neuronal theories, there has been a growing interest in using natural stimulation to observe the behaviour of neural systems. Historically, simple artificial stimuli that were easy to control and parameterize, such as light spots and gratings, were used in vision studies (Felsen and Dan, 2005; Simoncelli and

Olshausen, 2001). Although these traditional stimuli gave us a great insight into how the visual system works, they are not powerful enough to help us uncover the intricate details of perception.

Several response properties of neurons can become apparent only after employing natural sensory inputs as stimuli. In the early visual processing stream, feature sensitivity, sparseness, normalization models, and several other properties have been widely assessed using natural stimulation. However, the use of natural stimuli necessitates detailed consideration since these stimuli are highly complex and difficult to parameterize. Moreover, there is no explicit definition of naturalness for different species, while the need for standardization and categorization of different stimuli is pertinent to this field of study. Overall, obtaining a reliable understanding of the processing in neural circuits requires (1) data on the response of sensory systems to natural stimuli and (2) knowledge of the characteristics of these natural stimuli (Felsen and Dan, 2005).

It has been shown that natural visual stimuli, such as natural images and movies, drive an extended network of areas throughout a large portion of the cortex (Bartels and Zeki, 2005; Betti et al., 2013; Mantini et al., 2012; Rao et al., 2007; Sun et al., 2012; Tjan et al., 2006). Bartels and Zeki (2005) propose a method for segregating different visual areas via the measurement of BOLD signals while subjects were asked to view a James Bond movie. They reported that each visual area elicits a unique time course of activity during natural viewing, and thus, can be used as a fingerprint for identification (Bartels and Zeki, 2005). Similarly, another study combined EEG and fMRI recordings to reliably localize stimulus-driven activity during the viewing of natural movies (Whittingstall et al., 2010). In another fMRI study with similar conditions, the authors observed consistent interaction patterns among relevant brain regions during viewing of

natural videos. They also found a correspondence between the network of these regions and the structural connectivity pattern derived from Diffusion Tensor Imaging (DTI) (Sun et al., 2012). Further, another study used natural movies to investigate the sensory-driven networks of activity in human and monkey (Mantini et al., 2012).

As previously mentioned, the neural coding process appears to have adapted to the statistics of the natural environment. Higher degree of spatial correlation in natural images induces a sparser code in neural populations in which a smaller set of neurons is active at a given time. Several studies have reported this sparser activation in response to natural scenes (Dumoulin et al., 2008; Froudarakis et al., 2014; Rainer et al., 2001; Simoncelli and Olshausen, 2001; Vinje and Gallant, 2000). The degree of correlation in time can be assessed by looking for repetitive patterns in the power spectrum of the signal. An strong degree of self-similarity can also be observed in the power spectrum of natural sounds (Attias et al., 1998). In a behavioural study, the authors have shown that the amount of spatial coherence in the images is highly correlated with the subjective level of naturalness (Groen et al., 2013). Thus, spatial coherence indexes the amount of scene fragments (scene clutter) and has been observed to modulate the amplitude of the single trial evoked potentials from 100ms up to 250ms post-stimulation (Groen et al., 2013).

In addition, LGN neurons are known to whiten the spectrum of the visual stimulus. However, a whitened image still contains obvious structures that cause statistical dependencies among the receptive field responses of the simple cells. These dependencies mainly occur in locations representing abrupt changes in the luminance (i.e. spatially extended edges and contours) (Simoncelli and Olshausen, 2001). In an early study, researchers measured the amount of lateral

inhibition needed to cancel out this correlation and found a reasonable match with the inhibitory surround field measured from the interneurons in the eye of the fly (Srinivasan et al., 1982).

Apart from sparseness, other characteristics of neuronal processing have been shown to be consistent with natural stimulation. Felsen et al. (2005) recorded the sensitivity of single neurons in cat visual cortex to simple Gabor-like visual features. They observed change in this sensitivity when the features were embedded in a natural scene compared to random stimuli. They reported that complex cells responses are vigorously higher when their preferred feature is enclosed in a natural scene (Felsen et al., 2005). Another study has shown that the contrast histogram of natural images can be well fitted with a Weibull function using different beta and gamma values for different images (Scholte et al., 2009). They also reported that these two parameters explain up to 71% percent of early ERP variance of an EEG sensor overlaying the early visual cortex (Scholte et al., 2009).

Fourier transform has been used extensively in studying the characteristics of natural images. It has been shown that, in these images, the power spectrum falls at a frequency of $\frac{1}{f^2}$, where 'f' refers to the spatial frequency. Some studies report better performance in a discrimination task for images with a $\frac{1}{f^2}$ power spectrum and impaired performance for images that are less natural (Knill and Field., 1990; Parraga et al., 2000).

Compared to the power spectrum, the phase spectrum contains greater perceptual information and is known to be more important for neural sensory coding. Indeed, it is the phase spectrum that substantially influences the appearance and representation of an image (Felsen et al., 2005; Hansen and Hess, 2007; Henriksson et al., 2009; Thomson et al., 2000). However, the phase

spectrum is a complex feature of natural images and the direct relation between an image's phase spectrum and the corresponding perceptual processing of that image is largely unknown.

When the phases of different spatial frequencies become locally aligned, edges emerge in the image (Hansen and Hess, 2007; Henriksson et al., 2009) (Fig. 1-2). Edges contain high concentration of contrast energy and induce a sparse filter response. In other words, a sparser response of the Gabor filters to natural images arises from the contour structure in these images (Di Russo et al., 2012). Several studies attempted to assess the sensitivity of the brain's visual processing to the phase of natural images (Froudarakis et al., 2014; Hansen and Hess, 2007; Henriksson et al., 2009; Hubner et al., 1988; Rainer et al., 2002; Rieger et al., 2013).The randomization of the phase content preserves an image's power spectrum and global properties but changes its local properties. More specifically, first- and second-order statistics are maintained while higher-order statistics are removed (Dumoulin et al., 2008; Dumoulin and Hess, 2006; Froudarakis et al., 2014; Thomson et al., 2000).

A natural image has peaked local-contrast histogram and is sparser than a phase-scrambled noisy image (Dumoulin et al., 2008). An early study reported that the ability to discriminate a phase-randomized image from its original version increases monotonically with the level of phase randomization (Hubner et al., 1988). In another study, the authors investigated the amount of phase alignment needed at different spatial frequencies for a human subject to perform a scene-matching task. They reported that the extent of phase alignment required depends on the degree of structural sparseness in the scene (Hansen and Hess, 2007). Others observed an increasing selectivity to phase congruency in grating stimuli, from V1 to higher visual areas (Henriksson et al., 2009).

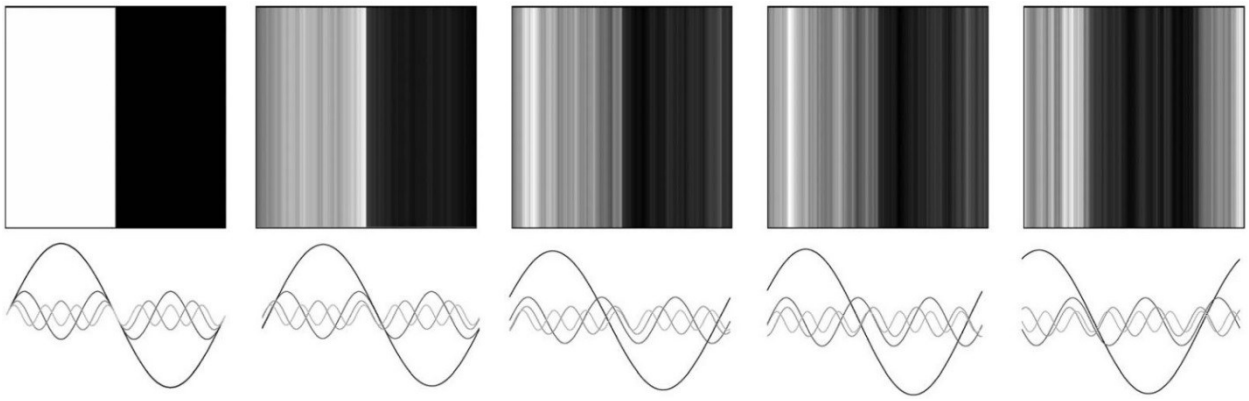

Fig. 1-2: the effect of phase scrambling on the edge representation. The top row represents a sharp edge manipulated by different levels of phase randomization. The bottom row shows the dependency of the edge salience on the local alignment of phases of different spatial frequencies (Hansen and Hess, 2007).

Human visual sensitivity to changes in the phase domain can be explained with fourth-order image statistics, such as skewness and kurtosis. These measures must be appropriately normalized to make them independent from the form of power spectra and eliminate the effect of second-order statistics. The normalization of these measures can be achieved via the flattening of the power spectrum (whitening) before computing higher-order statistics. The scales obtained through this process are called "phase-only skewness" and "phase-only kurtosis". These normalized measures convey a certain level of consistency among images, such that the "phase-only kurtosis" is always positive for natural images and close to zero for noise (Thomson et al., 2000). Importantly, these measures are highly correlated with observer threshold. In this behavioural study, the authors demonstrate that skewness and kurtosis are two measures that have non-overlapping histograms of natural and random-phase images.

1.5. V1 and higher areas responses to natural stimuli

Neural response characteristics at early stages of cortical visual processing, i.e. in area V1, are well matched with the properties of natural images (Olshausen and Field, 1996, 1997). V1 can be considered as a collection of local, oriented, and band-passed filters, whose response to natural stimuli demonstrates the sparseness property previously described (Dumoulin et al., 2008; Olshausen, 2013; Simoncelli and Olshausen, 2001; Vinje and Gallant, 2000). According to Olshausen and Field, the membrane potentials of V1 neurons are determined by the integration of feed-forward receptive-field response, as well as the short- and long-range feedback connections between cortical areas (Olshausen, 2013). For instance, a pattern context effect is defined by the modification of a neuron's response to a particular stimulus, which could be either suppressed or enhanced based on the coherence or dissimilarity of surrounding elements, respectively (Murray et al., 2002).

As mentioned above, cortical areas responsible for shape and object recognition are located in the temporal lobe of the brain (Larsson and Heeger, 2006; Malach et al., 1995). Recent studies have shown that LO activation increases when individual features in the visual input form a coherent object or shape. LO and other areas in the object recognition path, such as V4v and V3A, show decreased activation as the input becomes increasingly spatially scrambled and perceptually unrecognizable (Chen et al., 2010; Dumoulin and Hess, 2006; Grill-Spector et al., 1998; Malach et al., 1995; Murray et al., 2002; Rainer et al., 2002).

Several recent neuroimaging studies have investigated the neuronal activation of both primary visual cortex and higher-order areas in response to complex stimuli, such as objects, faces, and natural scenes. The level of recognisability or the degree of structural coherence can be

considered as the main variable in these studies. Based on the conventional notion of hierarchical processing stages, visual processing at the level of V1 is local and limited to basic image features, such as orientation and spatial frequency of edges. Meanwhile, perception and recognition of shapes and objects is believed to take place in higher-order visual areas. However, several recent studies proclaim that the traditional view of primary visual cortex as an inflexible module for low-level stimulus encoding is an oversimplification of the visual processing pathway. Most of these studies suggest that complex features of visual input modulate neural activity within V1 (Dumoulin and Hess, 2006; Henriksson et al., 2009; Murray et al., 2002; Rainer et al., 2001, 2002), denoting V1 as an adaptive and dynamic processor of information.

In a study by Murray et al. (2002), the degree of structure in the stimulus was altered by scrambling the edges of a shape or impairing the perception of form from motion in a set of moving random dots. Similarly, in Dumoulin and Hess (2006) study, circular shapes and flow-field patterns were formed from local Gabor filters. The coherence level of these units was then changed gradually. Both studies concluded that the BOLD signal in area V1 decreases as the input forms a coherent structure (Dumoulin et al., 2008; Murray et al., 2002). In another study, the authors observed increase in V1 BOLD signal upon the modification and camouflage of the face stimuli to a less recognizable form. Incomplete faces in this study elicited weaker BOLD signal in the fusiform face area (Chen et al., 2010). Moreover, BOLD signal from area LO and V3A are shown to increase with increasing shape coherence or decreasing image scrambling (Dumoulin and Hess, 2006; Murray et al., 2002; Rainer et al., 2002). A combined MEG/fMRI showed that the neural response in V1 increases for unrecognizable and masked objects from 67ms post-stimulation (Bar, 2003). In an fMRI study, the authors degraded natural images of

faces and bodies by gradually adding increasing levels of noise. They observed an increase in V1 activity along with a decrease in LO activity as noise increased (Malach et al., 1995). Other studies recorded the BOLD signal of different areas in an anesthetised monkey in response to various levels of image phase-scrambling (Rainer et al., 2001, 2002). They observed a non-monotonic change in the BOLD signal as a function of phase coherence in all investigated visual areas. They related this pattern of response to simultaneous changes in the number of active neuronal populations and the overall strength of activation in each population (Rainer et al., 2001, 2002).

Other studies failed to identify the modulatory effects of V1 activity by complex, global features of visual input. In an electrophysiology study on mice, the authors reported that, under both anaesthesia and active wakefulness, the overall activity in V1 does not differ significantly between natural and phase-scrambled movie viewing. However, in quiet wakeful mice, the activation is higher for the phase-scrambled condition (Froudarakis et al., 2014). Similarly, another electrophysiology study in monkeys revealed that area V2 responds more vigorously to naturalistic textures after 100ms, whereas the neuronal response in area V1 is similar to natural stimuli and their spectrally matched noise version. These researchers also obtained similar results using fMRI data in humans (Freeman et al., 2013). Another study used images of natural scenes under different levels of phase randomization. The authors concluded that the population response of V1 neurons, as measured by both fMRI and MEG, does not depend on phase perturbation (Rieger et al., 2013).

In Tjan et al. (2006) study, the research team investigated the sensitivity of different visual areas to the level of noise in the input image. They reported that the BOLD signal in area V3a and the

posterior fusiform sulcus (pfs) shows the strongest modulation by noise, whereas no noticeable effect can be observed in areas V1 and V2 (Tjan et al., 2006). In accordance with Tjan et al. study, another fMRI study concluded that the major variance of V1 activation (about 87%) can be explained by the level of total contrast-energy in the input, whereas extra-striate areas are more sensitive to sparsely distributed contours (78% of the variance in area V3) irrespective of the amount of overall contrast (Dumoulin et al., 2008).

In summary, primary visual cortex (V1) has been accepted as an active processing unit which gets input via both feedforward and feedback connections within the highly connected hierarchy of visual areas in the brain. This connection has been developed during an evolutionary process to adopt to the statistical properties of the environmental sensory inputs. In other words, our perceptual inference about the natural word is actively getting more efficient by learning from the natural environment and the characteristics of the expected sensory inputs. Natural images are considered as a useful stimuli to study this behaviour in the visual processing system. It has been shown that altering the naturalness of the visual input modulates the activation in different visual areas in the brain. The majority of relevant studies have reported stronger activation in higher-order object-sensitive areas as the natural stimuli becomes more recognizable. However, the existence of top-down modulation of early visual processing upon the perturbation of high-level stimulus features is still a matter of debate.

1.6. Research objectives

In this piece of work, I hypothesize that despite the low sensitivity of V1 neurons to higher-order features, which are mainly embedded in phase properties of natural sensory input, the population response characteristics of neurons in this area is different before and after removing the phase information from the natural stimuli. This difference in response can be interpreted to some extent by relying on proposed neural coding theories, which take into account the effect of feedback in sensory input processing.

In order to test this hypothesis, I progressively randomized the phase content of natural images while keeping the RMS contrast and the mean luminance constant. I presented the intact and phase-randomized images in a random order while recording the neural activity of the brain using magnetoencephalography (MEG). The high temporal resolution of MEG enabled the investigation of dynamic cortical responses in visual areas under differing conditions.

The remainder of this thesis describes the methods and results of this study, followed by a discussion and conclusion section. Firstly, the method section describes the stimulus design and the experimental paradigm in detail. Further, two pilot studies are briefly discussed. In addition, the process of MEG data acquisition and analysis is fully explained within this section. Secondly, the results section includes the main findings, experimental results, and statistical tests. Thirdly, the discussion section interprets the results and conveys the significance of these findings. Lastly, the conclusion states the main findings of this study, while postulating about their implications in future works.

2. Method

2.1. Stimulus design

2.1.1. Image preparation

I acquired 16 images of different animals in natural scenes from the Stock.xchng free image database (http://www.freeimages.com/). In each of these scenes, the animal's body was the dominant object, encompassing more than 70% of either the height or width of the image. The size and focus on the animal's body ensures that higher-order object-sensitive cortical areas, such as V3A and LO, are activated by the visual stimuli.

Once images were selected from the database, I converted them to gray-scale. I then resized their height and weight to 910 and 790 pixels, respectively. Another important step in preparing images for these kind of stimuli is to equalize the mean luminance and Root Mean Square (RMS) values of these images. This equalization step first required a non-linear, exponential increase of the intensity values of the pixels, which was performed using a constant power of gamma (1.8). The justification and significance of this exponentiation will be explained in detail in Section 2.1.5 (Screen Gamma Calibration). The mean luminance and the RMS values for a particular image are defined by Equation 2.1 and Equation 2.2, respectively.

$$\mu = \frac{1}{NM} \sum_{x=1}^{N} \sum_{y=1}^{M} p(x,y) \tag{2.1}$$

$$RMS = \sqrt{\frac{1}{NM} \sum_{x=1}^{N} \sum_{y=1}^{M} \left(p(x,y) - \mu \right)^2} \tag{2.2}$$

In these equations, N and M refer to the number of pixels in the height and width of an image and p(x, y) refers to the intensity value of the pixel in the position [x, y]. According to Equation 2.2, the RMS value can justifiably be considered as a measure of image contrast.

In order to set the mean luminance and RMS values of all the images to a single desired value, I applied Equation 2.3 to each image in my database.

$$I_{eq} = (\frac{RMS_d}{RMS_I}(I - \mu_I)) + \mu_d \quad (2.3)$$

In this equation, I refers to the intensity matrix of the original image while RMS_I and μ_I refer to the RMS and mean luminance values corresponding to the original image. RMS_d and μ_d refer to the desired values for RMS and mean luminance, respectively. The final image, I_{eq}, will have the mean luminance and RMS values equal to the desired values. In my stimuli, I chose μ_d to equal 78 and RMS_d to equal 32. It is important to note that the luminance of the gray background of the visual stimulus was set to equal the mean luminance of the images.

2.1.2. Phase scrambling

In order to obtain the phase-scrambled version of the natural images, I perturbed the phase content of the Fourier transform of each image. As shown in the schematic in Fig. 2-1, after computing the Fourier transform of each image, a random phase value was added to the original phase value. The random phase value was drawn from a uniform distribution in the range [-a, a]. In the next stage, I used the original Fourier magnitude and the manipulated phase content to compute the inverse Fourier transform and obtain the phase-scrambled version of the image.

It is important to emphasize that following the inverse Fourier transformation, some pixel values may exceed the acceptable range of image intensity values. In turn, these pixel values must be clipped to the attainable range in order to prevent luminance saturation or overflow. More specifically, there is a trade-off between the number of clipped pixels and the contrast of the image. I observed that as the contrast increases, the inverse Fourier transform following the phase scrambling results in more 'out of range' pixels.

Rieger et al. (2013) reduced the contrast and obtained less than 0.5% clipped pixels. In my stimulus, I reduced the RMS value (= 32) and the mean luminance (= 78) in order to assure less than 1% clipped pixels. This 'small pixel value' manipulation introduces negligible variance in the predefined mean luminance and RMS values. In particular, within my database of original and scrambled visual stimuli, the mean luminance and RMS values of all images remained within the range of 78±0.07 and 32±0.2, respectively. More specifically, fully scrambled images had a mean luminance of 78.01±0.04 and a RMS value of 31.8±0.1. Halfway scrambled images had a mean luminance of 77.99±0.04 and a RMS value of 31.9±0.1. Finally, intact images had a mean luminance 77.95±0.03 and a RMS value of 32.00±0.03.

The level of phase perturbation could be adjusted by changing the width of the interval from which the random phases are drawn. For instance, drawing a random variable from an interval between -180 to +180 degrees results in a completely randomized phase content similar to the phase content of white noise. However, if we use a shorter interval and draw the random value from -90 to +90 degrees, the reconstructed image would partially preserve edges (Fig. 2-1).

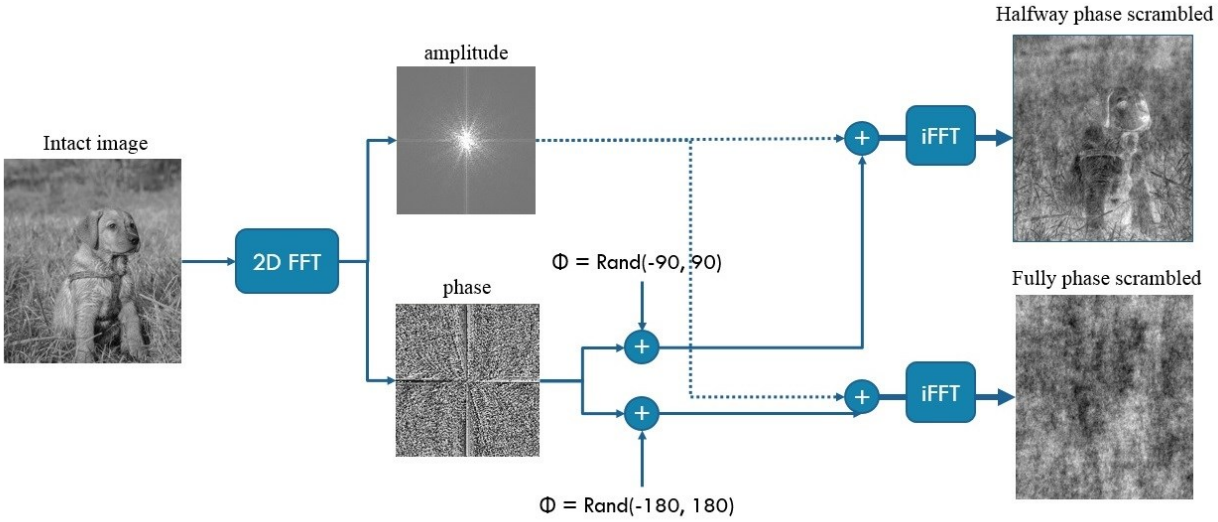

Fig. 2-1: schematic of the phase scrambling approach. The intact images on the left was decomposed to amplitude and phase content by applying 2D Fourier transform. The phase content was then manipulated by adding random phase values. In one level of phase scrambling, the random values were drawn from a uniform distribution in the range of -90° to 90° and in the other level the values were drawn from the range of -180° to 180°. The phase content obtained by this scrambling method was then combined with the intact amplitude content by applying the inverse Fourier transform. This results in phase-scrambled images. The image at the top right shows the result of the first level phase scrambling, therefore it still contains some edge information. The bottom right image is generated with the second level of phase scrambling which randomizes all the phase information and removes existing edges.

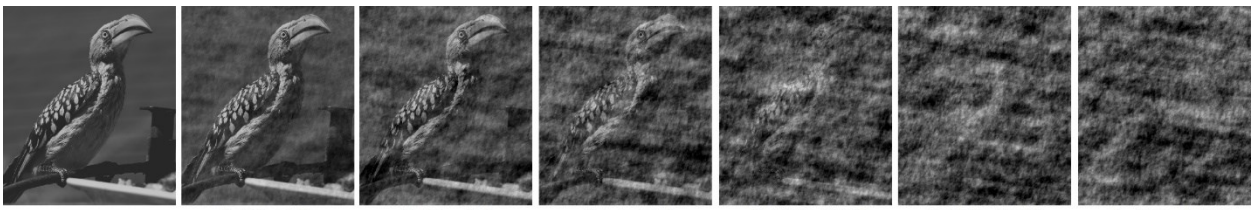

Fig. 2-2: Different levels of phase scrambling. The additive random phases were chose from a uniform distribution with the ranges indicated below each image (in degrees). Phase scrambling perturbs local contrast and distributes it over the image.

Fig. 2-2 shows gradual stages of the randomization of the phase content of a particular image. As previously mentioned in the introduction, manipulating the phase content alters higher-order statistics of the image but leaves the first- and second- order statistics unchanged (Dumoulin et al., 2008; Dumoulin and Hess, 2006; Froudarakis et al., 2014; Thomson et al., 2000). In order to verify this characteristic, I computed the phase-only kurtosis of my database images after

applying different levels of phase scrambling. Similar to the results in Rieger et al., phase-only kurtosis was computed following image whitening using Equation 2.4. The effect of whitening on a natural image is shown in Fig. 2-3.

$$k = \frac{\frac{1}{NM-1}\sum_{x=1}^{N}\sum_{y=1}^{M}\left(p(x,y)-\mu\right)^4}{\left(\frac{1}{NM-1}\sum_{x=1}^{N}\sum_{y=1}^{M}\left(p(x,y)-\mu\right)^2\right)^2} - 3 \qquad (2.4)$$

In Equation 2.4, N and M refer to the image dimensions and μ refers to the average intensity values. The subtraction of the value 3 enables the acquisition of the zero kurtosis value for Normal distribution.

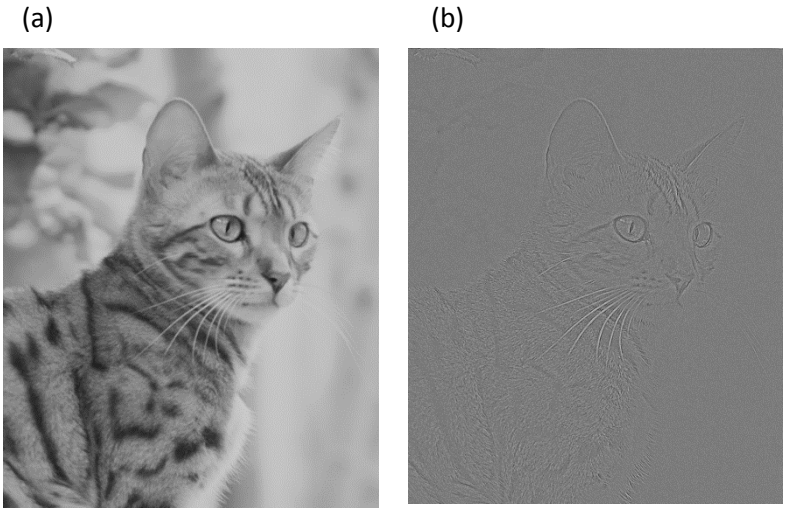

Fig. 2-3: The effect of whitening on an image. (a) a natural image and (b) the whitened version.

Fig. 2-4 shows the effect of phase scrambling on the second- and a fourth-order statistic: RMS contrast and kurtosis, respectively. The results are consistent with the expected positive kurtosis for natural images and the close to zero values for noise (Thomson et al., 2000).

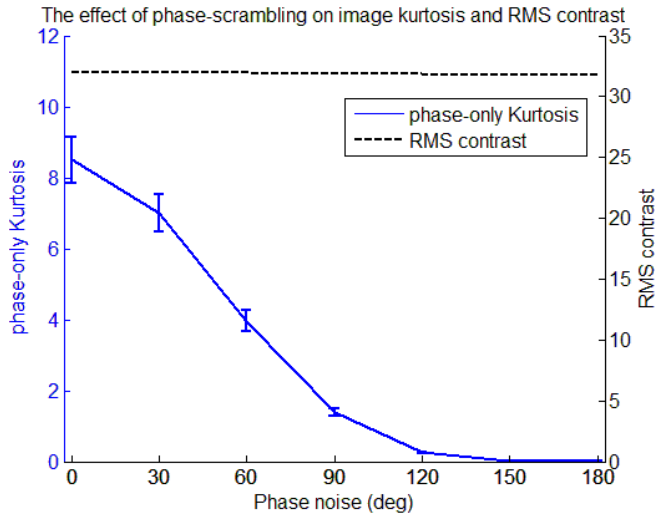

Fig. 2-4: The effect of phase-scrambling on the second and fourth order statistics of natural images. The horizontal axis indicates different levels of phase scrambling as shown in Fig. 2-2. The blue solid curve is the average of phase-only kurtosis computed over 16 natural images in my database. The error bars show the standard error of the mean over these 16 samples. The black dashed line corresponds to the average RMS contrast values. As the curve shows, these values remain approximately unchanged with phase-scrambling.

In my experiment, I employed only two levels of phase randomization for each image: completely phase scrambling, i.e. by choosing random values from -180 to +180 degrees, and halfway phase scrambling, i.e. by obtaining random values from -90 to +90 degrees (Fig. 2-1). Along with the intact version of each image, I prepared three conditions for each of the 16 images in my database, resulting in a total of 48 different images.

2.1.3. Edge smoothing

The borders between the image and the gray background potentially trigger activity in V1, since the neurons in this area respond vigorously to sharp edges and corners. To avoid this confounding effect, the edges of the images were smoothed with a transparent mask, which was based on a Gaussian cumulative distribution function. The mask overlays the image and alters the brightness of each pixel. This level of brightness is dependent on the pixel's distance from the image borders (Fig. 2-5).

In order to create the mask, I first defined an inner border, whereby all pixels within this border were unmodified by the smoothing process. The width of the mask border was chosen to be 20 pixels (0.38° of visual angle). Next, the distance between the inner border and each pixel external to this border was computed, demonstrating quantitatively the proximity of each pixel to the edge of the image. Then, for each pixel, this distance was used to calculate the level of manipulation and masking required via Equation 2.5.

$$C_i = \frac{1}{2}\left(1 + \mathrm{erf}\left(\frac{d_i - \tau}{\sqrt{2\sigma^2}}\right)\right) \tag{2.5}$$

In this equation, C_i is the coefficient of the mask for the ith pixel. The C_i values are all in the range of 0 to 1. d_i refers to the distance of the pixel i from the inner border. τ and σ are the parameters of the error function and were set to 0.5 and 0.15, respectively. τ shifts the center of the mask and σ indicates the slope of the filter. Fig. 2-5(a) shows the diagram of the mask function using these parameters. Since the d_i values were normalized, the horizontal axis includes values ranging from 0 to 1. The effect of this edge smoothing method on a natural image is illustrated in Fig. 2-5(b) and in Fig. 2-5(c).

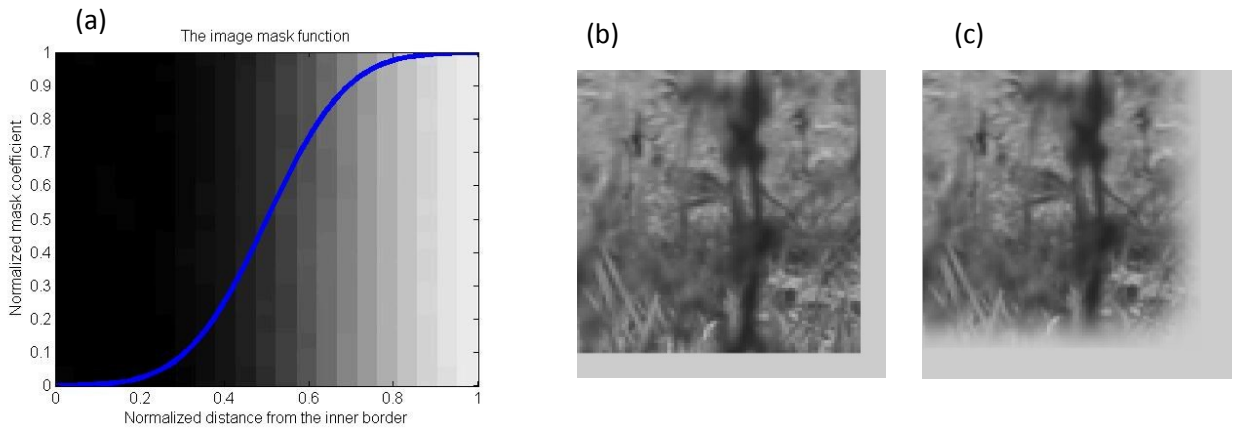

Fig. 2-5: Smoothing the edges of the images. (a) The transparent mask function used for smoothing the edges has the form of the Gaussian cumulative distribution function. The horizontal axes shows the distance of each pixel from the inner border of the image, i.e. as the distance increases, the pixel is more distant from the inner border and closer to the edge. Distance one corresponds to pixels on the edge of the images and distance zero is assigned to all pixels within right on the inner border. These pixels are not masked. The vertical axes represents the mask coefficient corresponding to each pixel. Pixels with the mask coefficient of one were completely covered with the mask: they were assigned the value of the gray background. (b) and (c) show a corner of a natural image before and after applying the smoothing mask, respectively. Following edge smoothing (c), the transition of the image to background is smooth and does not create sharp edges as in (b).

2.1.4. Stimulus presentation

The stimulus was designed using Psychtoolbox in MATLAB. Images extended over an area of 15° (width) by 17.25° (height) of visual angle. The image borders were 2.5° away from the fixation point along both the horizontal and vertical dimensions (Fig. 2-6(b)). As illustrated in Fig. 2-6(a), the images were presented on the left and right side of the lower visual field in a random order. Stimulus presentation was performed in these specific locations of visual space for four reasons.

Firstly, the primary visual cortex in humans is located in and around the calcarine sulcus. Therefore, the pyramidal cells in the dorsal and ventral parts of V1 are oriented in opposite directions. According to this specific anatomical property, stimulating half of the visual field (either upper or lower part) will result in a higher amplitude response from the bulk of neuronal

population in this area (Rauss et al., 2011). If both the upper and lower visual fields were stimulated, then the neural responses from the dorsal and ventral banks of the calcarine could potentially cancel each other out (washed out).

Secondly, an important control parameter for this study, the lower-order statistics of the image, is preserved when presenting the visual stimulus to only one half of the visual field (left or right). The left part of the visual scene is represented in the right hemisphere of the brain and vice versa. Therefore, if we were employed full-field images, only half of image was projected in either brain hemispheres. However, the phase-scrambling method employed in this experiment guarantees the equality of the global frequency content in the intact and scrambled images. Therefore, the frequency content of the right half of an intact image is not necessarily equal to the frequency content of the same half of the scrambled version. If the stimuli were presented over the left and right visual hemi-fields simultaneously, the frequency content of the projected image on each hemisphere could not be equally maintained between the natural and phase scrambled conditions. Thus, presenting images in either the right or left hemi-field is crucial for maintaining the control factor (frequency content) in this experiment.

Thirdly, keeping the stimulus away from the central vision helps binding the attention away from the images. In this experiment, we insist on reducing the effect of complex perceptual factors such as attention so that the obtained neuronal responses could be more inferable. .

Lastly, if the stimuli were presented more centrally, this position may stimulate a greater degree of noise. This increased level of noise would be expected since the magnification factor decreases with increasing eccentricity. Further, due to imprecisions in fixation, stimuli that are presented more peripherally will activate approximately consistent cortical regions over different trials, more so than stimuli presented centrally.

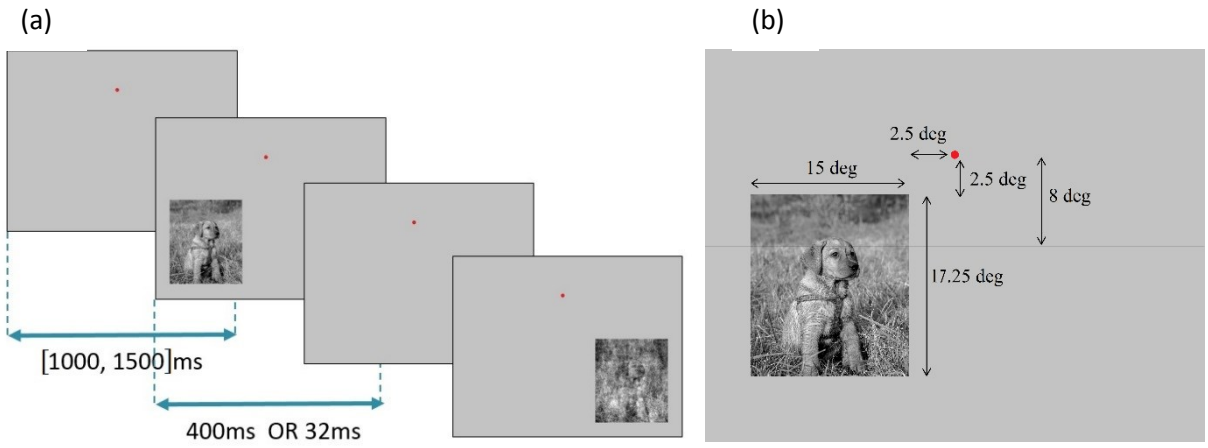

Fig. 2-6: The stimulus presentation. (a) Images were presented in the right or left lower visual field in a random order. The duration of image presentation was either 34 or 400ms and the inter-stimulus-intervals extended randomly within the range of 1 to 1.5s. Throughout the recordings, subjects had to fixate on the red dot and report any change in its color by pressing a button. (b) The exact location and extent of the images with respect to the fixation point and the center of the screen (values are in degrees of visual angle).

In addition to constrained image presentation I also added a second condition, named Null condition, which consists of only a gray background with no image presentation. This condition was presented in the exact same manner as any other image condition and was also used as a baseline condition for the statistical tests.

Throughout both 'image' and 'no image' presentations, a red fixation point was presented on the gray background to maintain subject fixation. The fixation point diameter was 0.38°. The color of the fixation point changed from darker to brighter red and vice versa, at random intervals between 0.8 to 3 seconds. In order to bind the subjects' attention away from the images presented, subjects indicate when the fixation point changed color by pressing a button. If the subject responded correctly within 0.5s after the change, the fixation point blinked as a sign of correct response.

Between the presentation of images and the Null condition there were inter-stimulus intervals (ISI) consisting of a gray background and a fixation point. The duration of these intervals was random, within the range of 1 to 1.5 second. This randomness prevented the subjects from predicting when the upcoming image would appear. In addition, lack of expectation reduced rhythmic neural stimulation and corresponding brain oscillations.

Images were presented either for a 'short' or 'long' time period. The 'short' duration consisted of 34ms, whereby the neuronal processing of the visual stimulus only occurs once the stimulus is no longer in the visual field (i.e. in the absence of any visual input). In contrast, the 'long' duration consisted of 400ms, enabling neuronal processing to occur in the presence of visual input. By comparing the brain activity under these two conditions, feed-forward and feedback effects could potentially be distinguished from overall neuronal activation.

Each of the 48 images was presented a total of 8 times: 4 times on the right visual field, 4 times on the left visual field, 4 times with 'short' duration, and 4 times with 'long' duration. In other words, within each run, an image was presented in a specific location and for a particular duration twice. In each MEG data-acquisition session, five runs were obtained, each taking approximately 11 minutes.

Overall, there were thirteen different conditions in the experiment: three levels of image scrambling, two hemi-field stimulations, two durations of presentation, and the Null condition (Fig. 2-7). In total, 160 trials per condition per MEG session were obtained (16 different images × 2 repetition in each run × 5 runs), and were usually reduced to about 140 trials following data de-noising.

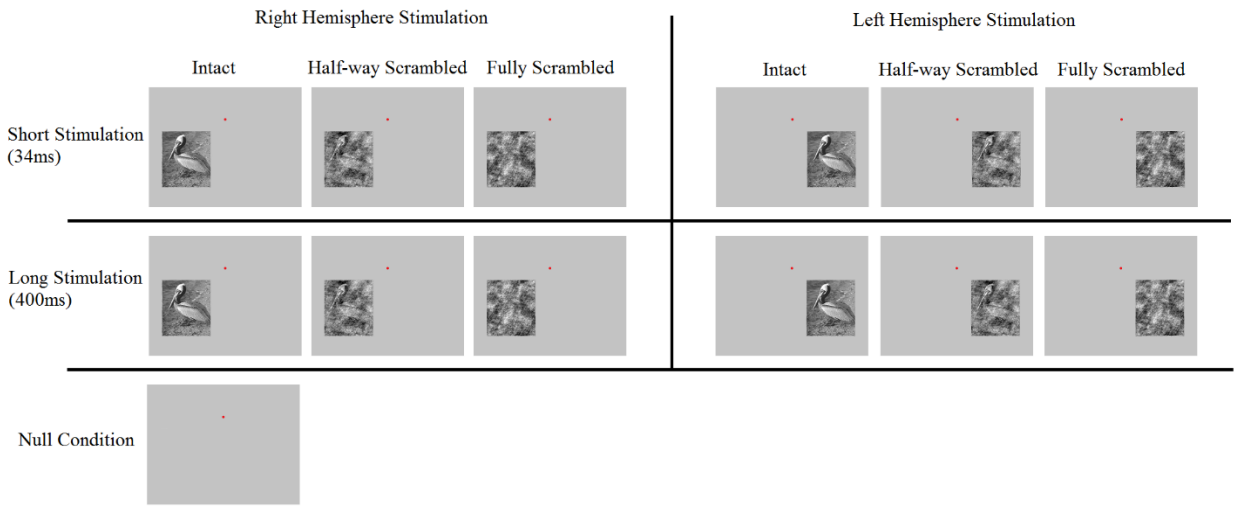

Fig. 2-7: Thirteen different conditions were included in the stimuli. Three types of images (intact, half-way scrambled and fully scrambled) were presented on the left and right hemi-filed and with short (34ms) and long (400ms) durations. The Null condition consisted of only a gray background with the red fixation point. This condition was used as a baseline for normalization and statistical testing.

2.1.5. Screen Gamma Calibration

In digital cameras, the relationship between the actual luminance received and the intensity value saved as an image is not linear. Using the non-linear transformation in Fig. 2-8(a), cameras can reserve more bits of memory for representing tones to which our eyes are naturally sensitive. This process is called gamma encoding and is defined by Equation 2.6. As shown in Fig. 2-8(a), the gamma value (γ) in the encoding process is lower than 1.

$$V_{out} = V_{in}^{\gamma} \tag{2.6}$$

On the other hand, this non-linear transformation goes in the reverse direction when the image is displayed on a monitor and the luminance of the natural scene is converted back into light. This latter process is called gamma decoding (Fig. 2-8(b)) and its gamma value is higher than 1.

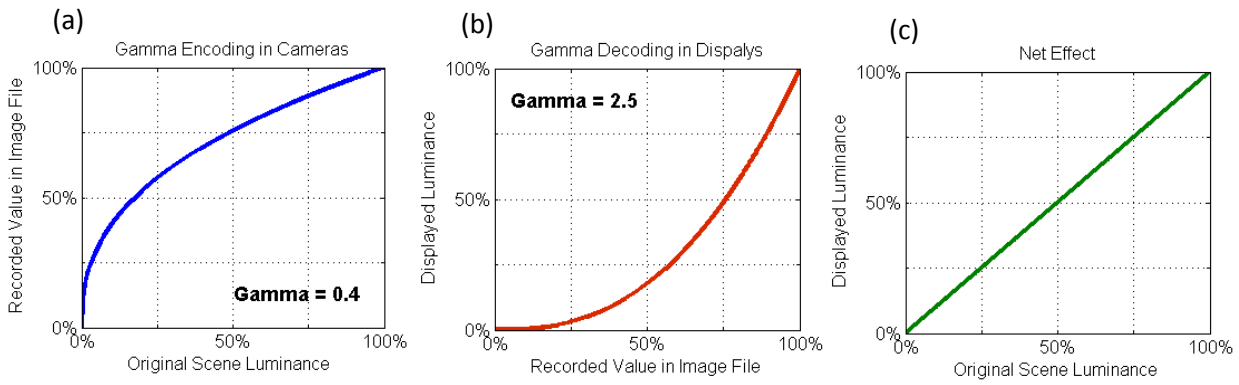

Fig. 2-8: Gamma encoding and decoding processes. The non-linear transformation from the original scene luminance to the recorded values in cameras (a), known as Gamma Encoding. An inverse transformation takes place when the image is displayed on a monitor (b), known as Gamma Decoding. The net effect of these two transformations is linear (c). Note that the Gamma values in Gamma Encoding and Decoding processes are smaller and larger than one, respectively.

Gamma decoding, also known as gamma correction, is a display property that should be considered carefully in vision studies. Through its non-linear transformation properties, gamma correction can seriously impair results. With respect to the visual stimuli used in this study, the intact images and their phase-scrambled version have exactly the same frequency content and approximately the same mean luminance and RMS values. However, if these images go through a non-linear transformation (the gamma decoding of the display), their frequency content will no longer be identical. In order to solve this problem, the transformation of intensity to luminance on the display must be linearized.

This linearization first requires the computation of the gamma curve of the display used in this experiment. In order to do so, a colorimeter (X-Rite DTP94) was attached to the display in the MEG room. During the gamma curve measurement, the screen was lit up with the three primary colors at varying intensities lit up the screen in a sequential order while the colorimeter measured the output luminance. The output of the colorimeter and the intensity values of the control

stimulus were then used to estimate the gamma curve of the display. The brightness, contrast, and gamma setting of the projector in the MEG room were adjusted in order to obtain the best estimation of the curve.

The final results are shown in Fig. 2-9. The blue channel showed an irregular behaviour, especially for high intensities, but this irregularity had an insignificant effect on the results. Based on the estimated curves (dotted line) in Fig. 2-9, the gamma value of the projector is approximately 1.8. In addition to the plotted graphs, a look-up table was computed from the inverse of the gamma curve (solid curve) in Fig. 2-9. This look-up table was later passed to the screen settings in the Psychtoolbox software to correct for the non-linear intensity-luminance transformation.

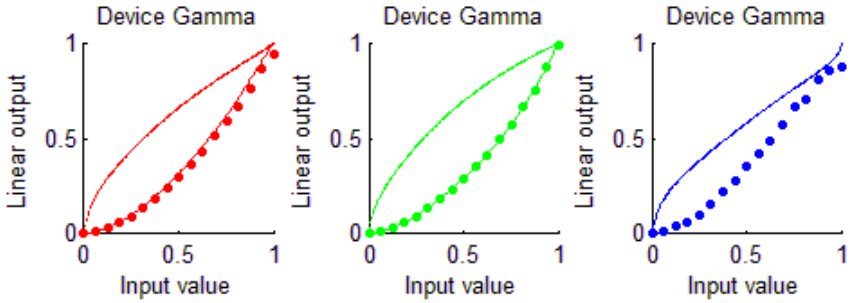

Fig. 2-9: Estimation of the projector gamma curve for the three color channels (dotted line). The intensity and contrast values of the projector were set to the default value which was 32 (max is 63). The gamma in the projector settings was set to maximum (15). This projector's configuration made the estimated transformation curve closer to a gamma function. The Solid lines show the inverse transformations used to correct for the nonlinear behavior of the projector.

As previously described, this process solves the problem of impaired frequency content. However, if one were to look at natural images on a calibrated screen, these images would appear reduced in contrast and strangely brighter due to the non-linear gamma decoding of these

images before reaching our eyes. Based on Fig. 2-8, the middle stage (b) was removed, which increases the mean luminance and reduces the contrast of the perceived images. In order to compensate for this change and recuperate the natural appearance of these images, a function (similar to Fig. 2-8(b)) was applied to the original images before the manipulation of the phase content occurred. For further explanation, please refer to Section 2.1.1, where I previously described the process of increasing the intensity values of pixels by exponentiation with a constant power (gamma = 1.8) (Fig. 2-8). Since this non-linear manipulation of the pixels was performed before the phase scrambling stage, the frequency content remained unchanged when displaying images. The manipulated images were presented on a corrected screen (i.e. with a linear intensity-luminance transformation function) and looked as an original image presented on a usual screen (i.e. with nonlinear gamma curve transformation) (Fig. 2-10).

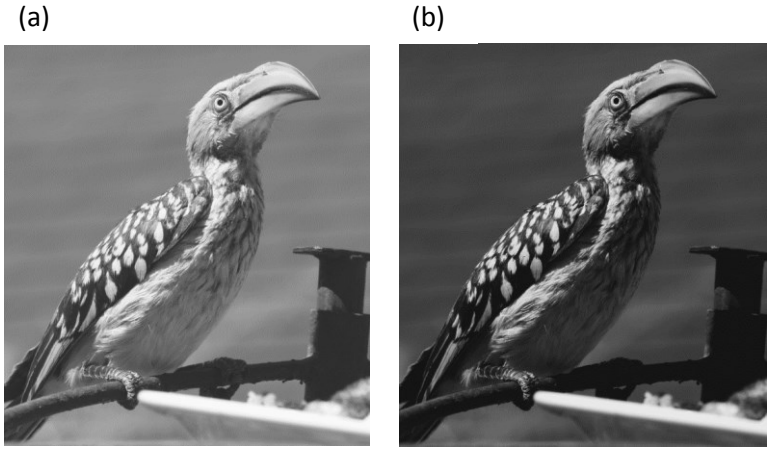

Fig. 2-10: Manipulating the intensity values of the images. A natural image before (a) and after (b) manipulating the intensity values. Image (b) on a display with corrected gamma curve looks like image (a) on a natural display with nonlinear intensity-luminance transformation.

2.2. Subjects

Eight subjects, three women and five men, participated in the main MEG session. Subject ages ranged from 22 to 45 years. Subjects had either normal or corrected-to-normal vision. For subjects with no previous MRI scan, a structural T1 scan was obtained prior or following the MEG session. Participants were required to answer the MRI safety questionnaire and change their clothes before entering the room. All subjects submitted written informed consent prior to the experiment. The experimental protocol was approved by the Ethics Committee of the Montreal Neurological Institute.

The same 8 subjects participated in a secondary MEG recording session, which used a checkerboard stimulus. In addition, two subjects participated in an fMRI retinotopy session and three subjects participated in a psychophysics session. Except for the MEG session using checkerboard, none of the subjects in the pilot studies were included in the main MEG study.

2.3. Pilot data acquisition and analysis

2.3.1. fMRI retinotopy

Retinotopy mapping, a special fMRI scan, is accepted to be the most precise means of delineating distinct visual areas. Several areas in the visual cortex are known to represent spatially ordered, continuous maps of the visual field. In other words, adjacent neuronal populations in each area represent corresponding adjacent receptive fields of the visual field. This preservation of adjacency can be used to specify the retinotopic boundaries between different visual areas. While different locations of the visual field are sequentially and

systematically stimulated, the Blood Oxygenation Level Dependent (BOLD) signal from the activated regions in cortex is recorded. Upon further analysis, the different locations in the visual field can be matched to the activations on the cortical surface. Further, each complete map can be specified as a single visual area.

Two commonly used visual stimuli were employed for retinotopy mapping: an expanding ring and a rotating wedge. The ring and wedge sweep the visual field at different eccentricities and polar angles, respectively, and elicit an ordered activation of the cortex. As shown in Fig. 2-11, the ring and wedge are represented by alternating patterns: black-and-white checkerboard, coloured checkerboard, or moving random dots. This patterning helps activate all visual areas, considering different areas are differentially sensitive to contrast, color, and motion.

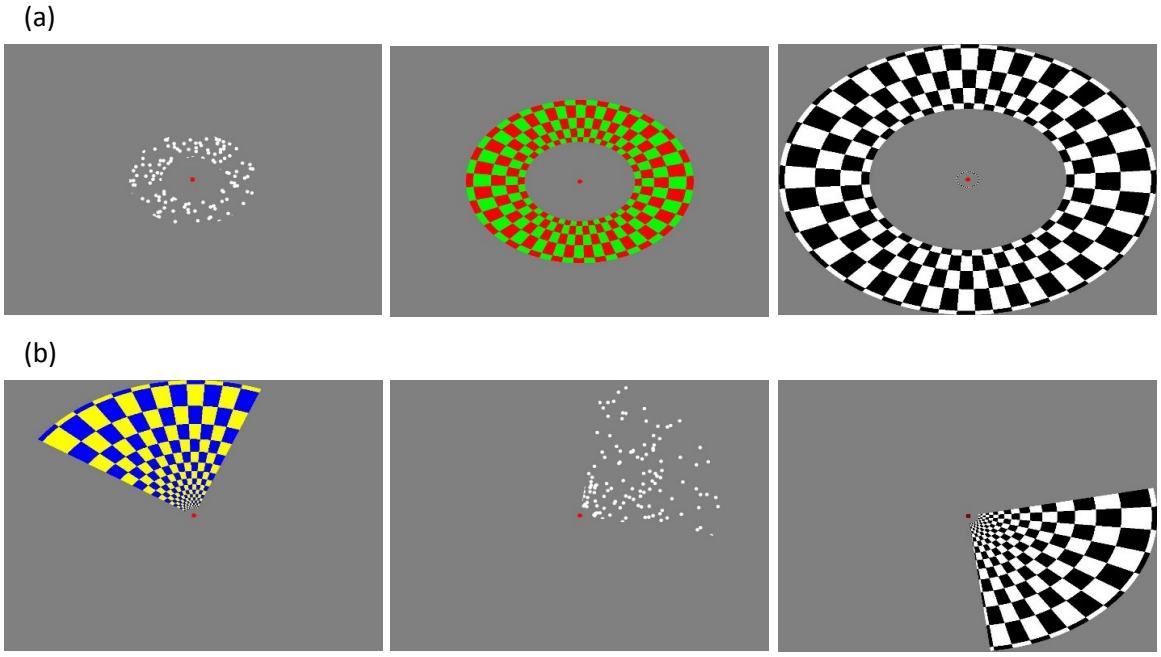

Fig. 2-11: Screenshots of the retinotopy stimuli. (a) Expanding and contracting rings and (b) rotating wedges, composed of checkerboard and random moving dots.

There are different areas in the visual pathway that show sensitivity to specific types of stimuli. For instance, area LO in the ventral pathway is sensitive to objects and shows strong response when viewing recognizable objects in natural scenes. Moreover, emphasizing motion (i.e. by contrasting moving and static dots) can activate area MT in the dorsal pathway. In order to localize these visual areas, their BOLD signals can be recorded while the subject is presented specifically designed visual stimuli (Fig. 2-12). Details regarding the stimulus parameters and fMRI data acquisition are provided in the Appendix 7.1.

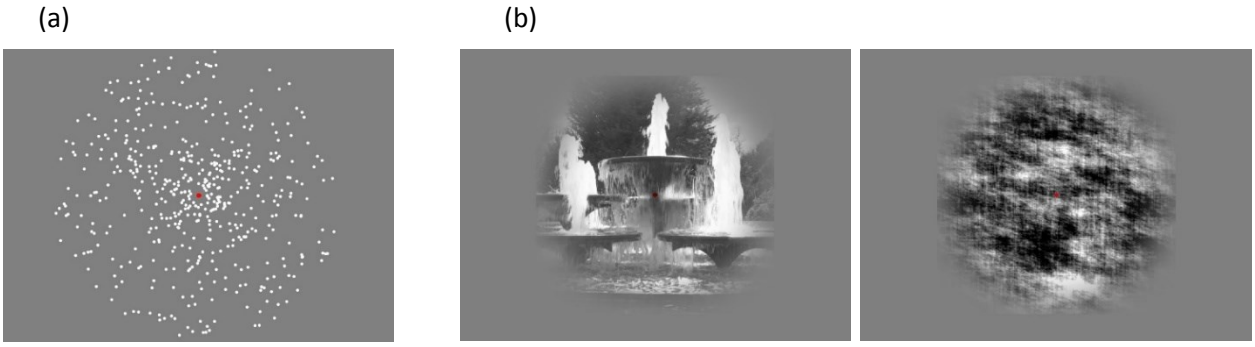

Fig. 2-12: fMRI visual stimuli for activating specific visual areas. (a) MT-localizer consists of a full field random dots which alternates between the static and moving conditions. (b) LO-localizer contrasts between images of natural objects and scenes (left) and their phase scrambled version (right). The edges of the images were smoothed using a method similar to the one explained in Fig. 2-5 but using a circular smoothing mask instead.

Although this approach is appropriate for delineating visual areas of interest, it was excluded from the first stage of this project for several reasons. Firstly, the main interest of this study is activation of V1, an area located in the upper and lower banks of the calcarine sulcus. Area V1 is relatively easy to delineate from the structural T1 data alone, and therefore, this approach is not necessary. Secondly, the high resolution of the BOLD signal and precise mapping of the visual areas can be fully appreciated when used as priors to guide the ill-posed source localization problem in MEG, which is not the approach employed in this experiment.

2.3.2. MEG scan using checkerboard stimuli

In addition to the conduction of two MEG pilot scans, which was used to adjust the experimental paradigm and details of the main stimuli, a standard visual stimulus was designed. This stimulus was checkerboard pattern, comprised of black and white checkers with full contrast on a gray background. The checkerboard stimulus was presented in exactly the same manner as the natural image stimuli (Fig. 2-13), for both 'short' (34ms) and 'long' (400ms) duration periods. Further, two types of checkerboard stimuli were shown, which were reversed in contrast (black quadrilaterals turn to white and vice versa).

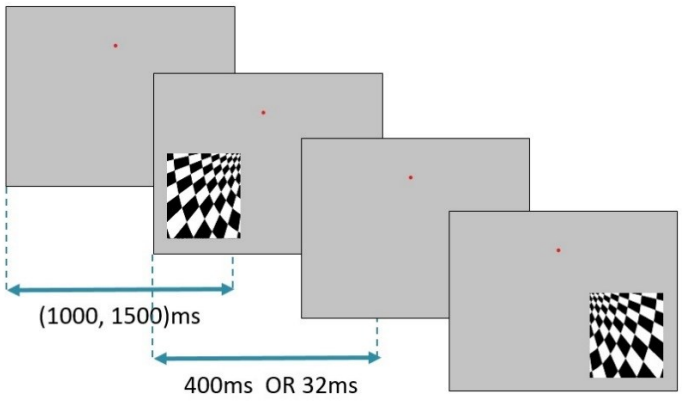

Fig. 2-13: Checkerboard stimulus presentation. Checkerboard stimulus was designed the same way as the main natural image stimulus (Fig. 2-6) by replacing images with black and white circular checkerboards.

The presentation of the standard checkerboard stimuli and the recording of corresponding neuronal activity served as a control. In other words, responses to these stimuli, termed standard Visual Evoked Potential (VEP) responses, were compared to responses triggered by the presentation of intact and scrambled images of natural scenes. This comparison enabled the identification and characterization of three distinct phases involved in visual processing for both the VEP and natural image responses. Since the shape of the VEP response varies across

individuals, using a standard stimulus designed for this specific purpose is expected to be helpful for capturing this specificity.

Two subjects were scanned in the MEG using this paradigm. There were four runs, each lasting approximately seven minutes, which resulted in 128 trials per condition. The data was analysed using the Brainstorm software. The analysis procedure was similar to the procedure employed for the main experimental data (see Section 2.4).

2.3.3. Psychophysics study

As previously described, subjects were asked to indicate when the fixation point changed color. This attention-binding task ensured that different levels of attention did not confound the neural signal recorded throughout stimulus presentation. However, it is conceivable that a given subject may focus on one condition of the natural image presentation more than the other. For instance, the viewer may deem intact images to be more interesting since these images contain a meaningful object, whereas the halfway phase scrambled images may provoke subject curiosity to uncover the underlying shapes and representations.

In order to ensure that the designed attention-binding task was equally effective for different image conditions, a behavioural pilot experiment was conducted. For this experiment, the stimulus was similar to the MEG stimulus and the experiment was run using the psychophysics setup in Dr. Curtis Baker's lab at the Royal Victoria Hospital. A Mac LCD monitor was used for presenting the stimuli. An Eyelink eye tracker system was employed to follow the position of the eye throughout the test. For the purpose of this study, it was necessary to assess participant performance for different stimulus conditions, since it is possible for subject to physically focus

on the fixation point while simultaneously attending to images presented in their periphery. The experiment setup designed for this study is shown in Fig. 2-14.

Three subjects (including the investigator) participated in this experiment and a total of five runs were acquired, each lasting 10 minutes. The analysis of this data was similar to the subject's response analysis explained in Section 2.4.2. The results of this study are provided in Section 3.2.2, along with the results from the response analysis of the main MEG study.

(a) (b)

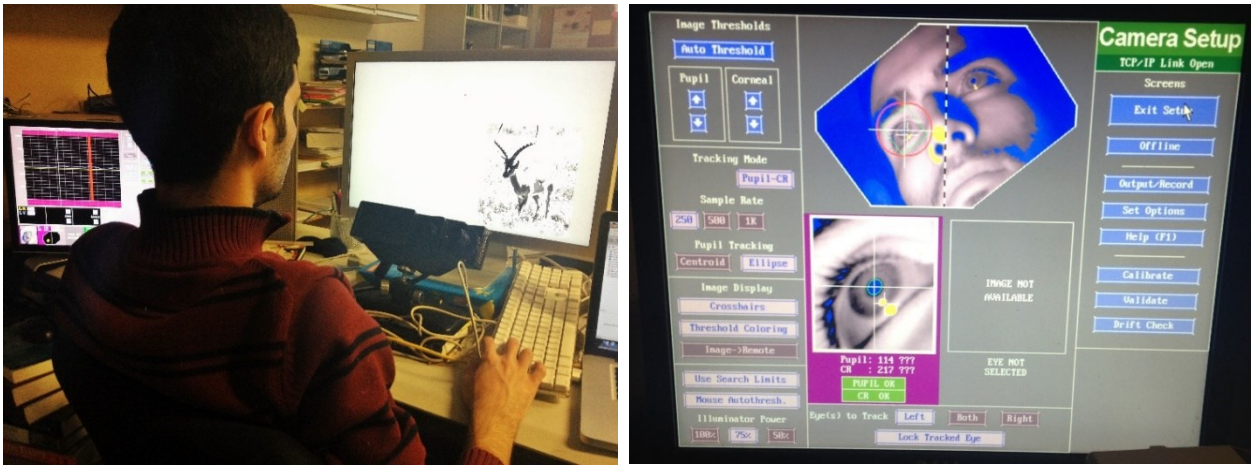

Fig. 2-14: (a) Pilot Psychophysics experiment setup. (b) The Eyelink eye tracker system used for this experiment.

2.4. MEG recording and data analysis

Electromagnetic brain activity was recorded using a whole-head 275 channel MEG system (CTF/VSM). The MEG signal was sampled at 2.4 KHz on all channels. The head positioning indicators were attached to the scalp and they monitored head movements during the session. EOG and ECG signals were acquired simultaneously in order to correct for the cardiac and eye blinking artifacts at the pre-processing stage. Further, an eye-tracker with a 148Hz-sampling rate

was used to record eye position and monitor subject fixation performance throughout the experiment. The eye-tracker was calibrated before each run. The empty room noise was obtained with and without the eye-tracker installation, ensuring that these signals did not interfere and taint the results. Two minutes of empty room (without a subject in the scanner) noise data was acquired for each session and was later used for computing the noise covariance.

A projector (Sanyo PLC-XP57L) was used to present the stimulus on the back-projection screen of the MEG lab, which was in front of the subject. The screen was 42cm×32cm in width and height, respectively, and was positioned 42cm from the subject's nasion. Subjects were asked to perform an attention-binding task, which required them to press a response button box with their index finger. Two button boxes were provided, for both of the subjects' hands, as subjects were instructed to alternate hands between runs. The use of left and right hands alternatively prevented lateralized motor activation in the brain.

Two markers were used to keep record of stimulus timing during data acquisition. The MATLAB program communicated with the MEG data acquisition system via an Analog port. When a new condition was initiated, either an image or blank screen, the program sent a trigger to the MEG system along with an 8 bit code that indicated the condition. Different codes for each of the 16 images and for each level of phase scrambling were chosen. In addition, different codes were assigned for the position (left or right hemi-field) and duration (short or long) of the image presentation.

Considering the random order in which the images were presented, this coding was necessary for the categorization of trials in our subsequent MEG analysis. However, it is important to note the 54±6ms delay between the execution of the command in MATLAB and the actual appearance of

the condition on the screen in the MEG lab. To keep record of this systematic delay, a small white circle of light was flashed at the beginning of each stimulus condition. A photodiode, which was attached to the left bottom corner of the screen, captured this flash of light and produced a pulse signal in response. This signal was detectable during data analysis and served as a marker. Information from both the Analog port and the photodiode was used to deduce the exact onset of each condition and to label the data accordingly.

In the following sub-sections, the analysis executed on different forms of acquired data is explained. Most statistical tests and analysis were performed using R. The use of MATLAB and Brainstorm is noted when applicable. Moreover, following any significant ANOVA result, the subsequent pairwise comparison tests were conducted using Tukey's honestly significant difference test in R. All p-values reported in the result section are two-tailed.

2.4.1. Eye-tracker data analysis

In order to analyse the eye-tracker data effectively, the recorded eye positions needed to be matched with both the stimulus onset and the recorded MEG data. In addition to creating triggers to mark these events in the MEG system, similar triggers were designed for the eye-tracker system. To mark the exact timing of stimulus onset, the photodiode information was needed.

Since the photodiode information was being sent to the MEG system, the synchronization of the eye-tracker and MEG systems was required. These systems were synchronized using an ongoing 1Hz pulse with duty cycle of 50%. The pulses were saved on both the MEG and eye-tracker systems along with the MEG data. In both systems, the last pulse preceding the initiation of a specific trigger was identified in order to match the timing of both systems. The difference between this time and the photodiode time was measured and added to the pulse time in the eye-

tracker data. In doing so, the actual event onset could be identified in the eye-tracker system (Fig. 2-15). The delay between eye-tracker triggers and the actual stimulus appearance on the screen was 42±5ms.

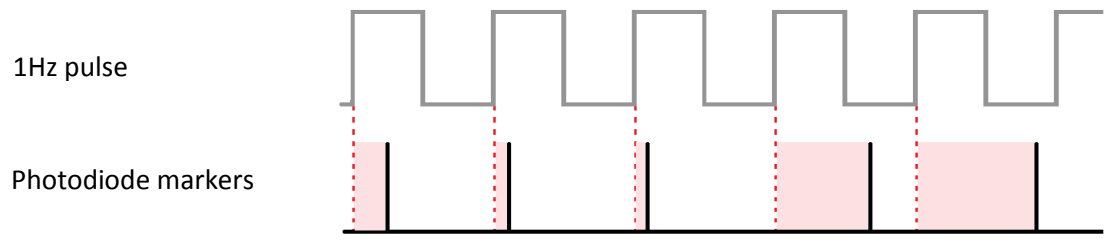

Fig. 2-15: Using 1Hz pulses to synchronize the MEG and eye-tracker systems. The difference between the photodiode marker (black bars in the bottom row) and the rising edge of the latest pulse is shown with shaded area. This difference is then added to the corresponding pulse time in the eye-tracker system to find the actual event onset in the eye-tracker data.

Having the exact time of events in the eye-tracker data enabled the division of eye position recordings into 1.3s epochs, extending from -300ms prior to stimulus onset until 1000ms post-stimulation. Thus, there were 44 eye-tracking samples in the baseline and 148 samples after the onset. Trials containing any missing value due to blinking or impaired pupil detection were removed from the data set. The eye-tracker data from subject 5 was excluded from further analysis due to a large percentage of missed points (about 65% of the entire data set). The remaining epochs were sorted into eight groups according to different stimulus conditions: fully scrambled, halfway scrambled, intact, and null condition, each of which presented for either a 'short' (34ms) or 'long' (400ms) duration.

A k-means clustering method, proposed by Koenig and Buffalo, was implemented to detect and separate eye movement and fixation periods. This method clusters the eye-tracker data based on

the eye movement distance, velocity, acceleration, and angular velocity. It also specifies the start and end of the saccades and fixations periods (Konig and Buffalo, 2014). The MATLAB source code for this method is provided on the website of BuffaloLab at University of Washington. Following the clustering of the eye-tracker data, a threshold of 1 degree was used to separate micro- and macro-saccadic movements based on Martinez-Conde et al., (2009). Trials lacking macro-saccadic movements were excluded from further eye-tracker data analysis. (Martinez-Conde et al., 2009).

For each time point in a given epoch, the percentage of trials containing macro-saccades in each condition was computed. The resulting eight signals per subject were derived, corresponding to eight different conditions. For each time point, a one-factor within-subjects Analysis of Variance (ANOVA) was conducted to reveal any significant difference in eye movement between trials of different image types, including the blank condition. To compensate for multiple comparison in the time domain, 10 consecutive p-values above the threshold (0.05) were required to be considered a significant difference (Forman et al., 1995; Rieger et al., 2013).

The standard deviation of eye position at each time point of a trial was then computed, acting as a quantitative measure of eye movement in each subject separately. This computation required the use of a moving window of 10 sample size (68ms), which measured the standard deviation of x and y components individually. These components were then summed together, resulting in a total standard deviation of eye position. For each time point, a one-factor within-subjects ANOVA was performed to find any significant difference in the standard deviation of eye movement between different conditions. Similar to previous analysis, to compensate for multiple

comparison, 10 consecutive p-values above the threshold (0.05) were required to be considered a significant difference (Forman et al., 1995; Rieger et al., 2013).

2.4.2. Subjects' response analysis

To ensure that there was no attention bias towards certain types of the images, the subjects' responses to the fixation task were analyzed. The first stage of analysis consisted of the categorization of changes in fixation point color based on time of occurrence. Next, changes that occurred during the blank period were separated from changes that occurred during stimulus presentation. Then, responses to changes in fixation color were assessed for each condition of stimulus presentation, including the identification of missed and delayed responses. A threshold of 0.5s was chosen for an acceptable response, resulting in an average of 70% correct responses among all subjects. A one-factor within-subjects ANOVA was performed to compare the percentages of correct response between different conditions. This analysis was conducted for both the overall responses of all subjects and individual subject responses between different runs.

In addition, response delays for each subject and each condition group were computed and plotted as a histogram. The Kolmogorov-Smirnov test was then performed to compare every pair of histograms corresponding to different conditions of the stimulus. Kolmogorov-Smirnov is a statistical test implemented in MATLAB, evaluating the null hypothesis that two sample sets are drawn from the same distribution.

2.4.3. MEG data pre-processing

Brainstorm, an open-source software application (http://neuroimage.usc.edu/brainstorm), was used for the analysis of the MEG data (Tadel et al., 2011). First, the data for each subject was processed separately. Then, a group analysis was conducted.

Individual cortical surfaces extracted from the MRI structural data was imported for each subject. These surfaces were then down sampled such that they consisted of 15,000 vertices in each source space. Prior to importing the raw MEG file for each run within a session, the empty room noise from that corresponding session was imported and its power spectrum density was computed. The signal characteristics in each MEG channel was monitored using the spectrum figure in Brainstorm. Any irregular pattern or peaks in the noise spectrum were inspected carefully.

In this manner, the malfunctioning channels could be easily detected and the interfering sinusoidal signals could be specified (Fig. 2-16). In some MEG sessions, one bad MEG channel was found and its signals were removed from the rest of analysis. The main interfering sinusoidal signals observed in other channels were the 60 Hz noise signal from the power line and its harmonics at 120 and 180 Hz. These harmonics were removed using the sinusoidal removal method in Brainstorm, which identifies the sinusoidal components and subtracts them from the signals in the time domain. Once interfering signals were removed, the noise covariance was computed from the empty room noise data. This estimation of the sensors' noise level was later used for the source reconstruction process (see Section 2.4.6).

With respect to the raw file of each run in the session, the power spectrum density was computed and the interfering sinusoidal signals were removed. A band-pass filter of 0.3 to 200Hz was used

to further filter the signal. In addition, the heartbeat and blinking artifacts were removed via the Signal Space Project (SSP) approach, a method implemented by Brainstorm.

Based on SSP method, the repetitions of the artifacts can be identified, either manually or automatically in Brainstorm, based on the recorded ECG and vertical EOG signals. The peak times of the QRS waves and the blinking bumps were saved for subsequent SSP calculation, in which a time window surrounding these artefact peaks is obtained. These time blocks are then concentrated into a matrix, from which the singular value decomposition can be computed. The singular vector with the highest singular value is chosen to calculate the projection operator. Finally, the method applies this projection onto the MEG data to remove the corresponding artifacts. After the SSP process is complete, a manual inspection was performed to mark the time epochs that contained muscle noise. The EMG noise is represented as bursts of high frequency signals over all the MEG sensors. These marked epochs were removed from further analysis.

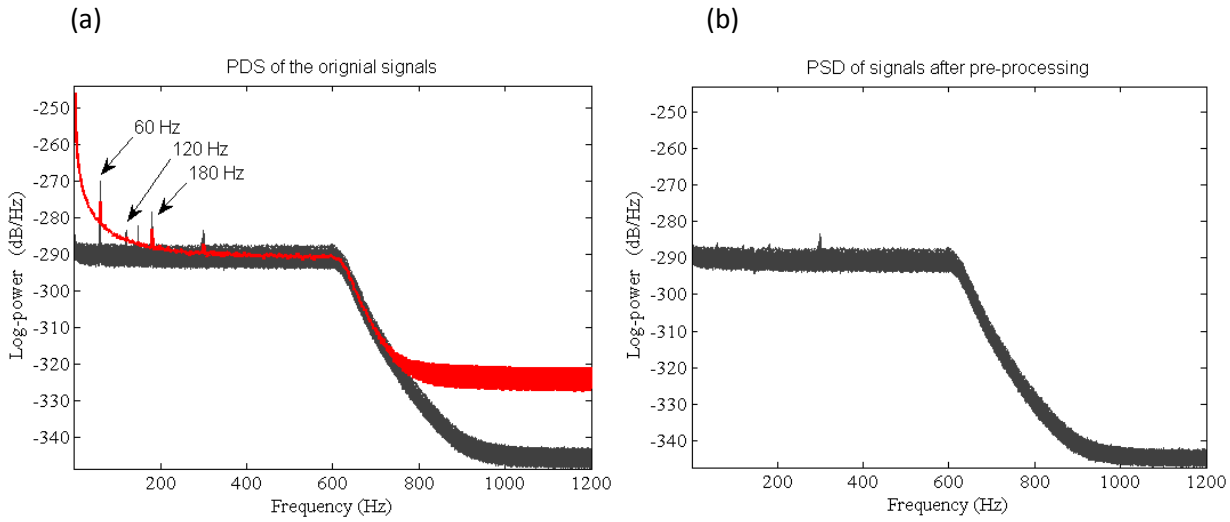

Fig. 2-16: The effect of removing bad channels and sinusoidal artifacts from the channels data. The Figure shows the Power Spectrum Density (PSD) for all MEG channels of the empty room noise data before (a) and after (b) pre-processing. The red channel in (a) does not function properly. The peaks in the power of 60Hz and its harmonics indicate the interference of the power line artifact. In (b) the bad channel has been removed and the sinusoidal artifacts have been subtracted using Brainstorm.

Following the cleaning of the data, it was important to organize the trials into distinct categories. Based on the stimulus paradigm of this study, there are thirteen defined groups (Fig. 2-7). One group included trials of the null condition, which were used as baseline for the normalization of data and the conduction of statistical analysis. The twelve remaining groups were divided according to their position within the visual field (right or left hemi-field), their level of phase scrambling (intact, halfway, or fully scrambled), and their presentation duration (short or long).

The onset of each trial was marked based on a simultaneously recorded photodiode signal. The MEG signal was then cut into 1.3-second epochs: from 300ms pre- to 1000ms post-trial onset. Every trial was corrected for DC offset with respect to its 300ms baseline by computing the average value across time (from -300ms to 0) for each channel and subtracting this average value from all values recorded in the trial. In order to remove large muscle artifacts from the data, trials that had at least one channel where the maximum peak-to-peak amplitude exceeded 3000 femto-Tesla (fT) were removed. Moreover, head movements were assessed using the head positioning indicators attached to the scalp and were monitored to ensure that they did not exceed 10mm of motion during each run. Nevertheless, more than 90% of total runs contain head movements of 6mm or less. These small head movements are negligible in terms of the following analysis.

2.4.4. The analysis of Visual Evoked potentials in EEG and MEG studies

Prior to describing the MEG data analysis in detail, the general characteristics of the Visual Evoked Potential (VEP) and basic concepts used to interpret this signal in MEG and EEG studies are explained. In the remainder of this study, these concepts are used to both propose a method

of analysis and interpret the results. The Visual Evoked Potential is the electrical potential recorded from the population of active neurons following visual stimulation. High contrast standard visual stimuli, such as gratings and checkerboards, are used to study the characteristics of this signal.

In EEG studies, three major components are generally detectable in a pattern on-set VEP. The first component, also known as the N75 or C1 component, peaks at approximately 70-90ms. This component has been extensively studied and is accepted to represent primary activation in the visual cortex. The second component, termed P100 or P1, is reversed in polarity and peaks at approximately 80- 100ms. The third component peaks from 120-180ms and is called the N145 or N1 component. It has been shown that the evoked response topography and waveform are highly variable across individuals (Ales et al., 2010; Hagler et al., 2009; Kelly et al., 2008; Rauss et al., 2011).

The VEP response and its components are regarded slightly different in MEG studies. Firstly, the peaks of the evoked response measured in MEG can be either positive or negative. Secondly, unlike EEG studied, the polarity does not have a functional interpretation. Thus, components are labeled with 'M' in MEG studies, rather than 'N' or 'P'. Thirdly, there is no direct equivalent of the C1 component in MEG studies, and in turn, most MEG studies fail to detect a prominent component of the VEP response before 100ms post-stimulation (Rauss et al., 2011). In general, two prominent neuromagnetic responses are detectable in response to visual stimuli. The first component, M100, arises at approximately 45-60ms and peaks at 100ms post-stimulation. The second component, M170, peaks at approximately 170ms.

2.4.5. MEG sensor-space analysis

In order to evaluate which sensors were responsive to the set of stimuli for each subject and each hemisphere, the trials of all conditions were averaged and the cumulative histograms of peak-to-peak sensors' amplitude were computed. For each individual, the number of sensors having at least 40% of the maximum peak-to-peak amplitude was determined. This value represented an average of 40 sensors, ranging from 31 to 45 sensors for different individuals. In the following analysis, 40 sensors (the average number over all subjects) having maximum peak-to-peak amplitudes were selected for each subject and were considered to be representative of active sensors.

Next, the average MEG signal from the trials of each condition was computed. The RMS value of the previously selected 40 sensors was calculated for every time point using 2.7:

$$RMS(t) = \sqrt{\frac{1}{N}\sum_{i=1}^{N} C_i(t)^2} \qquad (2.7)$$

In this equation, N refers to the number of selected sensors and $C_i(t)$ refers to the signal of the ith sensor.

Following this calculation, the RMS signals that corresponded to distinct stimulus conditions and time points were compared via a one-factor within-subject ANOVA. Multiple comparison in the time domain were compensated by the fact that only a minimum of 10 consecutive p-values above the threshold (0.05) was considered to be a significant difference (Forman et al., 1995; Rieger et al., 2013).

Moreover, the RMS values for specific time epochs were averaged, resulting in a single RMS value per condition for that given period of time (Fig. 2-17). Responses to short and long presentations of stimuli were divided into three time epochs, ranging from 30-70, 80-120, and 140-200ms post-stimulation. The first time period captures the early visual activation, known by EEG studies as the C1 component, which peaks at approximately 50ms post-stimulation. The second and third time periods were selected according to the two main VEP components in MEG studies, M100 and M170, respectively (Henson et al., 2003; Henson et al., 2007; Henson et al., 2009; Huberle and Lutzenberger, 2013; Meeren et al., 2013; Meeren et al., 2008; Tanskanen et al., 2005). With respect to neural responses triggered by 'long' (400ms) stimulus presentation, two additional time epochs were considered: 250-400ms and 450-510ms. These two time periods were selected based on the averaged response to the long duration stimuli (Fig. 2-17 (b) and (d)). The 250-400ms time window corresponds to steady state brain activation, which follows the evoked response and precedes the off-response. The 450-510ms time window captures the off-response, which peaks at approximately 480ms. Further, a one-factor within-subject ANOVA was performed to compare the average RMS values for each of these time epochs between different stimulus conditions.

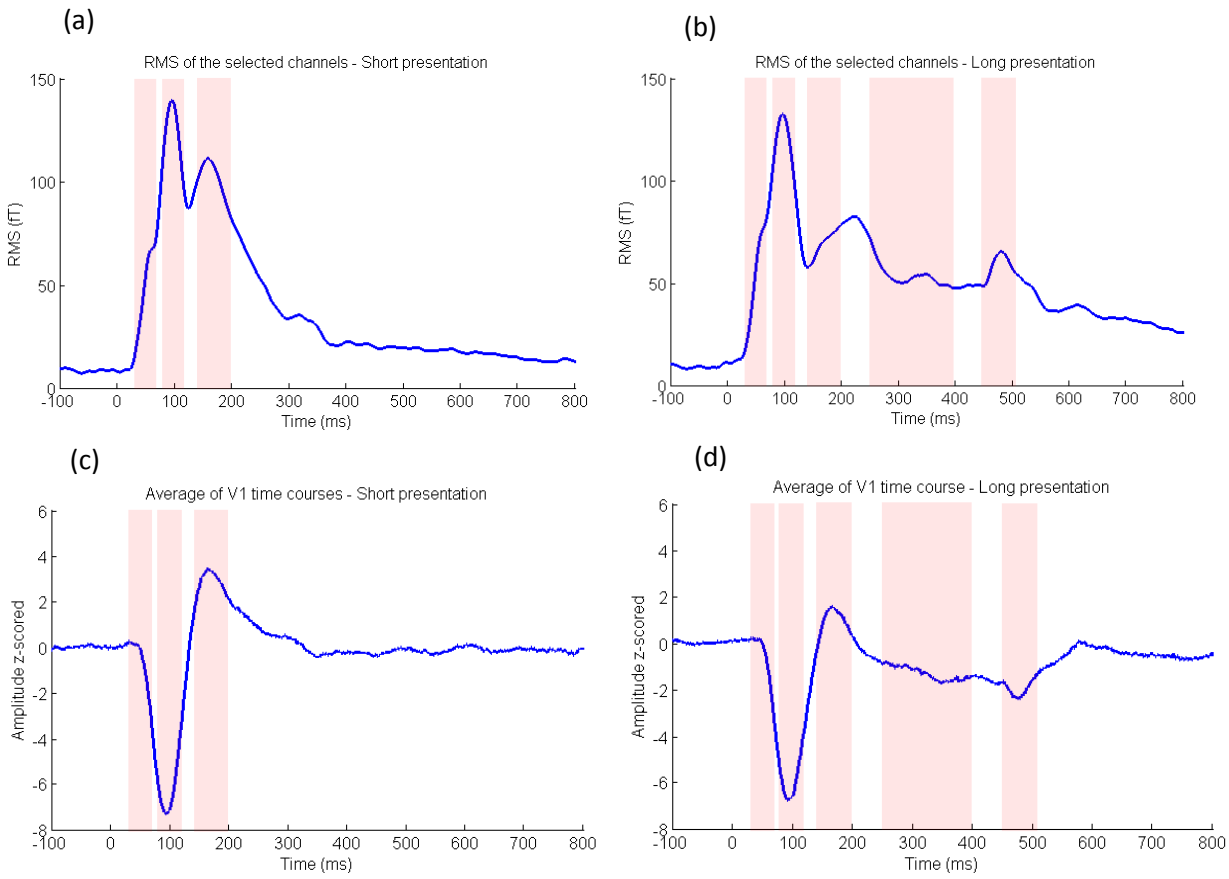

Fig. 2-17: The selected time epochs (shaded areas) used for statistical analysis in the sensor ((a), (b)) and source ((c), (d)) spaces. Plots show the average of response over stimulus condition, subjects and hemispheres. (a) and (b) show the RMS of 40 channels with maximum peak to peak amplitude. This 40 channels were selected for each subject separately. (c) and (d) are the VERs localize in area V1. (a) and (c) correspond to the short stimulus presentation (34ms) whereas (b) and (d) correspond to the long stimulation (400ms). The three first time epochs were the same for all signals: [30-70]ms captures the very early visual activation which is known as the C1 component in EEG studies (peaks around 50ms). [80-120]ms and [140-200]ms time epochs were selected according to the two main VEP components in MEG, M100 and M170 respectively. Two additional, late time periods were selected only for long presentation responses ((b) and (d)). [250-400]ms corresponds to the steady state brain activation following the evoked response and before the off-response. [450-510]ms captures the off-response which peaks around 480ms.

2.4.6. MEG source-space analysis

To estimate the sources of MEG signal on the cortical surface, a forward model is required. The overlapping sphere method implemented in Brainstorm was employed to obtain the lead-field matrix. Once a forward model and the noise covariance were attained, the Minimum Norm technique was used to solve the inverse problem and estimate the sources of MEG signal.

To define cortical regions of interest (ROIs), the sources maps from all conditions were averaged. The early visual activation from 30ms to 70ms post-stimulation was visualized according to these maps.

Based on previous studies, during this time period, the visual activation is dominated by area V1 (Rauss et al., 2011) Anatomical and functional cues were used to delineate this region. Since the stimulus was presented in the lower part of the visual field, the ROIs were constrained to the upper portion of the calcarine sulcus. The spatial extent was also adjusted based on the visual activation on the source map (Fig. 2-18). This process was performed manually for each subject, and in turn, anatomical differences among individuals could be evaluated.

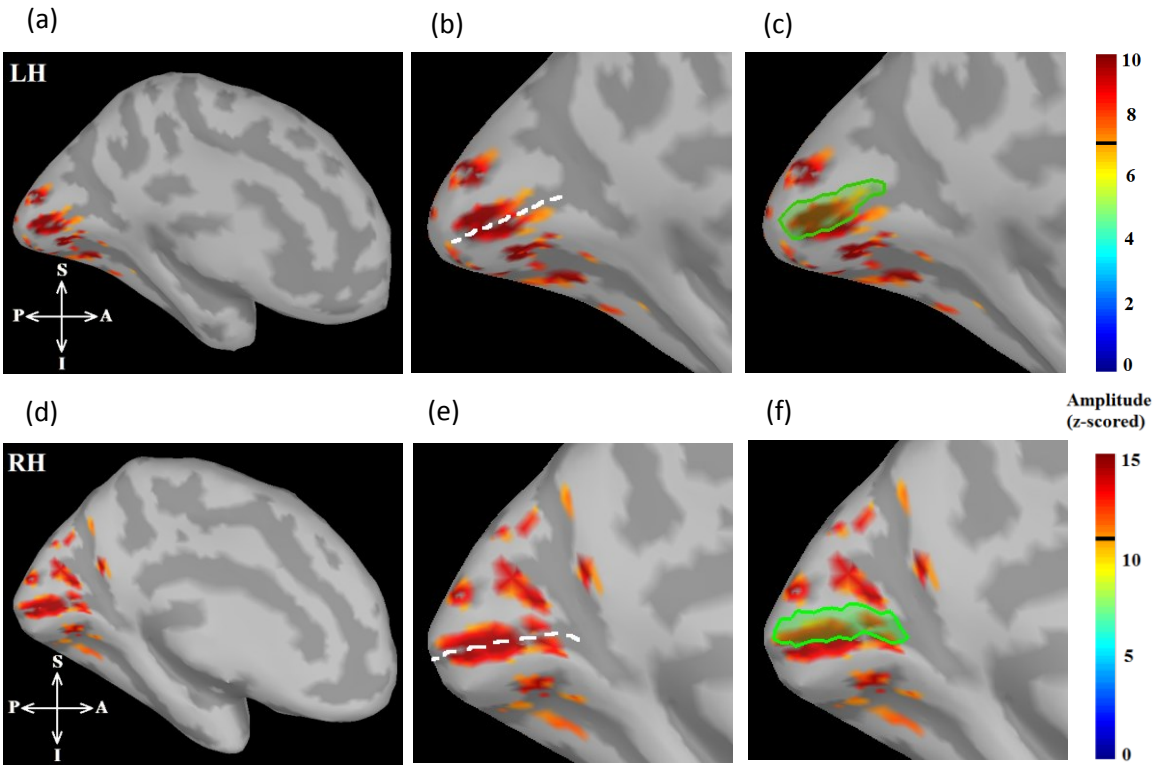

Fig. 2-18: Process of defining the region of interest in area V1. Figures show the brain activation for all conditions of the left (top panels) and right hemisphere stimulations (bottom panels) in one subject. The source maps are presented after averaging in time from 30ms to 70ms post-stimulation. Panel (a) and (d) show the sagittal view from inside of the brain of activation in the left (a) and right (d) hemispheres. Arrows indicate the Superior (S), Inferior (I), Anterior (A), and Posterior (P) directions. Panel (b) and (c) represent the magnification of the occipital cortex in (a).The calcarine sulcus is indicated with white dashed line in (b) and the selected ROI with green shaded area in (c). (e) and (f) represent the same thing as in (b) and (c), respectively, but in the right hemisphere.

In the next stage of processing, the mean response of sources within V1 was extracted for each hemisphere and each stimulus condition. A two-factor within-subjects ANOVA was performed, in which "image scrambling level" and "stimulated hemisphere" were the two within-subjects factors. The main interest of this test was on the simple main effect of "scrambling level" conditioned while there is no significant "scrambling level × hemisphere" interaction effect. As executed in previous statistical tests, only a minimum of 10 consecutive p-values above threshold (0.05) was considered to be a significant difference. In addition, similar two-factor within-

subjects ANOVA test was conducted with respect to the average values of the time epochs. In this case, the selection of the time epochs was identical to that of the sensor space analysis (Section 2.4.4). The time epochs are represented by shaded areas within Fig. 2-17(c) and (d).

As previously alluded to, the time-course shape of the Visual Evoked Response (VEP) varies between subjects. For instance, the early component of the VEP (known as C1 in EEG studies) may not necessarily be captured at similar times for different subjects. This variable time course is due to differences in cortical shape and orientation with respect to the sensors. The early peak (from 50-60ms post-stimulation) in the overall visual response was detected separately for each subject and aligned according to peak times. This process enhanced the clarity of the average VEP signal. A portion of the signal surrounding this peak was extracted. Further, a two-factor within-subjects ANOVA was performed on each time point and on the average of 30-70ms time interval values.

Aside from investigating the VEPs in the time domain, the time-frequency decomposition of V1 activation was also obtained. First, the time-frequency maps for each trial were computed using Morlet Wavelets, ranging from -300ms to 1000ms post-stimulation and representing frequencies less than 120Hz. The central frequency of the Morlet wavelet was 1Hz and the time resolution (FWHM) was set to 3. These maps were computed for each voxel in the ROI and were later averaged over all the voxels.

Further, for each stimulus condition, the time-frequency maps of all trials (both left and right hemisphere) were averaged to obtain the induced power of different frequency bands in time. To correct for the 1/f scaling in frequency powers, the z-score of the time-frequency maps was computed based on the power in the 200ms pre-stimulation period. Due to the boundary effect of

the time-frequency decomposition, the power of frequencies less than 16Hz cannot be identified for all the time points in the baseline. Thus, it is important to note that frequency bands less than 16Hz were excluded from the analysis. Finally, a paired t-test was performed using Brainstorm, which compared the time frequency maps of intact versus fully scrambled stimulus condition. The p-values were corrected for multiple comparison using False Discovery Rate (FDR).

3. Results

3.1. Pilot study results

3.1.1. fMRI retinotopy results

The retinotopic maps obtained from the fMRI data analysis of one subject are represented by Fig. 3-1. These maps were overlaid onto the cortical surface of the occipital lobe, which was sectioned along the calcarine sulcus and flattened. In the middle of the figure, the color wheels denote the corresponding visual field area that was stimulated. UVM, LVM, RHM, and LHM labels on the color wheels refer to the upper and lower vertical meridians, and the right and left horizontal meridians, respectively. Panel (a) portrays the polar angle map, which is a representation of neural response to the rotating wedge stimuli (Section 2.3.1). Panel (b) includes the eccentricity maps for the left (LH) and right (RH) hemispheres, which were obtained using the expanding and contracting rings stimuli (Section 2.3.1).

The vertical (solid white lines) and horizontal (dashed white lines) meridians could be delineated from the polar angle maps (Panel (a)) since the borders are represented by reversals in the polar gradient. For instance, in the dorsal portion of the left hemisphere (LH) of Panel (c), the transition from green to blue to red indicates the transition from the horizontal to the vertical meridian. Beyond this point, the gradient sign flips and the colors transition in reverse order, from red to blue to green, until the horizontal meridian is reached again.

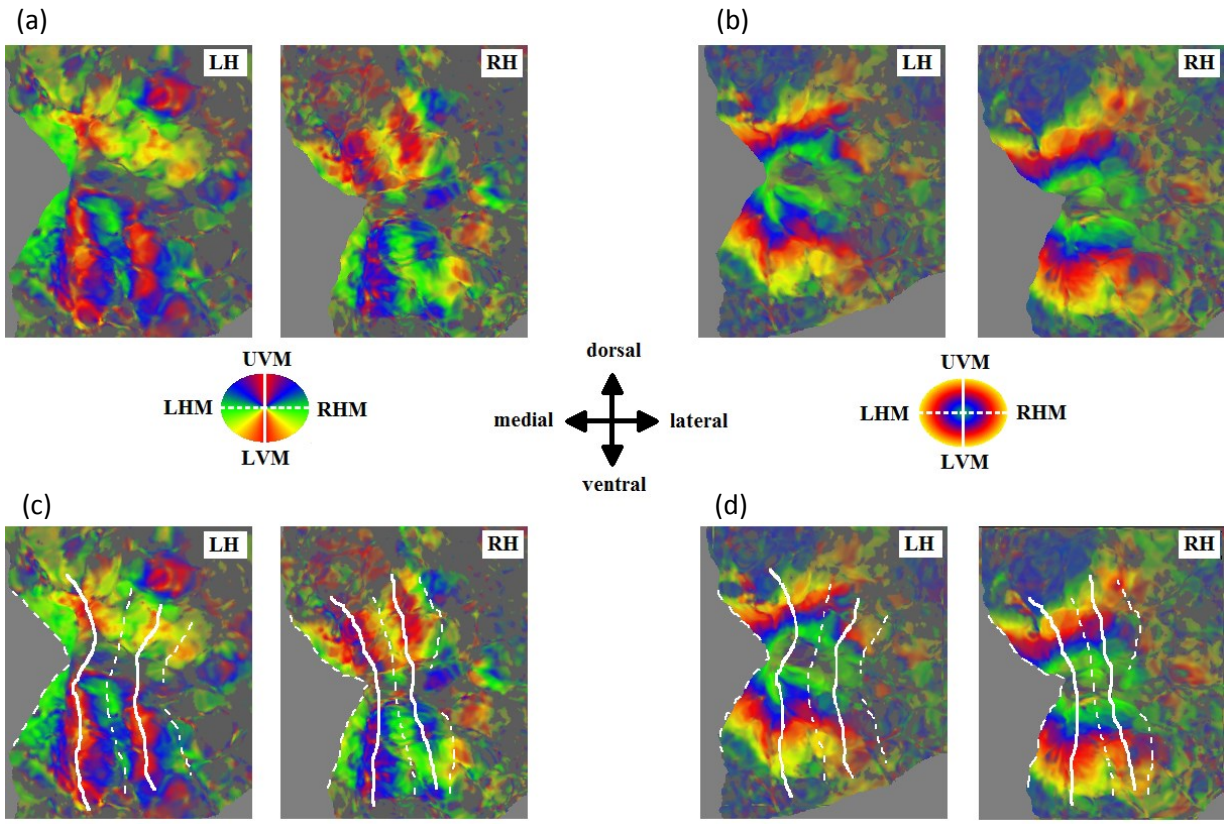

Fig. 3-1: The results of retinotopy analysis projected on the flattened surface of the occipital lobe. The cortical surface was cut along the calcarine sulcus (the most medial part of it, see the dashed line in the bottom figures). (a) Shows the map of the polar angle on the right (RH) and left (LH) hemisphere. The eccentricity maps are shown in panel (b). Panel (c) and (d) have the same retinotopic maps as in (a) and (b) respectively, overlaid with the vertical (solid lines) and horizontal (dashed lines) meridian markers. In the middle color wheels, UVM and LVM stands for upper and lower vertical meridian respectively, whereas LHM and RHM stands for left and right horizontal meridian.

A BOLD contrast analysis between dynamic and static conditions was performed using the MT-localizer stimulus (Section 2.3.1). The results for one subject are depicted in Fig. 3-2. Similarly to Fig. 3-1, the significance maps are overlaid onto representations of flattened cortex. Red areas indicate regions of positive contrast, i.e. the BOLD response to the dynamic condition (moving random dots) was greater than that of the static condition (motionless random dots). Blue areas depict regions of negative contrast, i.e. the BOLD response was stronger for the static condition.

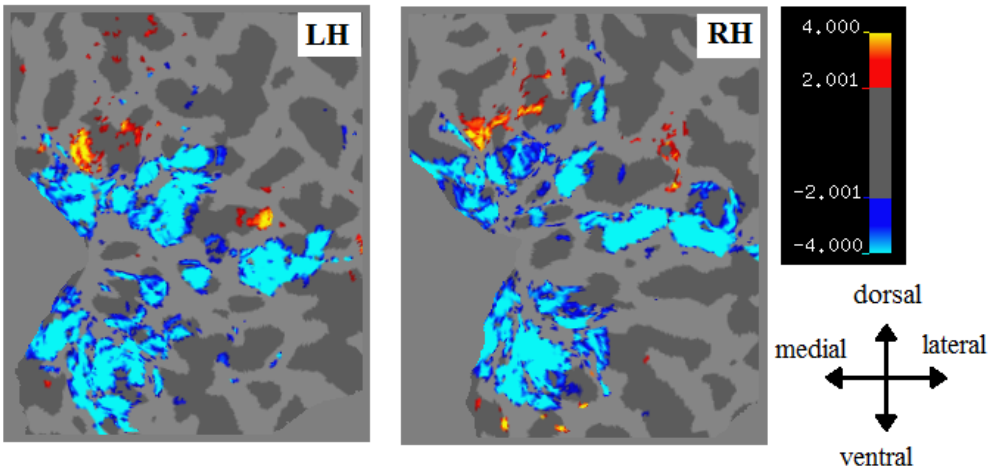

Fig. 3-2: The result of BOLD contrast analysis for the MT-localizer experiment overlaied on the flattened surface of the occipital lobe of the left (LH) and right (RH) hemispheres. Red areas indicate the regions which show higher BOLD response to moving dots compared to static dots. In contrast, the blue regions showed less sensitivity to moving dots compared to static dots.

3.1.2. MEG response to Checkerboard stimuli

Prior to presenting the results from the pilot MEG study that used checkerboard stimuli, it is important to look at Fig. 3-3, which is representative of the data obtained from the main MEG experiment that used natural images as visual stimuli. The time course of V1 neural responses for each subject, which was derived by averaging the response to all 'short' (34ms) conditions, is depicted in Fig. 3-3(a) and (b). The signals were then normalized using the 300ms pre-stimulus period. In section (c) of Fig. 3-3 the four panels correspond to the individual responses for four distinct subjects. As shown by this figure, the early component of VEP responses is clearly present for some subjects, even though early (less than 100ms) peaks in MEG evoked potential are expected to be absent (Rauss et al., 2011). Moreover, the typical M100 and M170, which are expected in MEG visual evoked responses, are not detectable in all of the time-courses presented in Fig. 3-3.

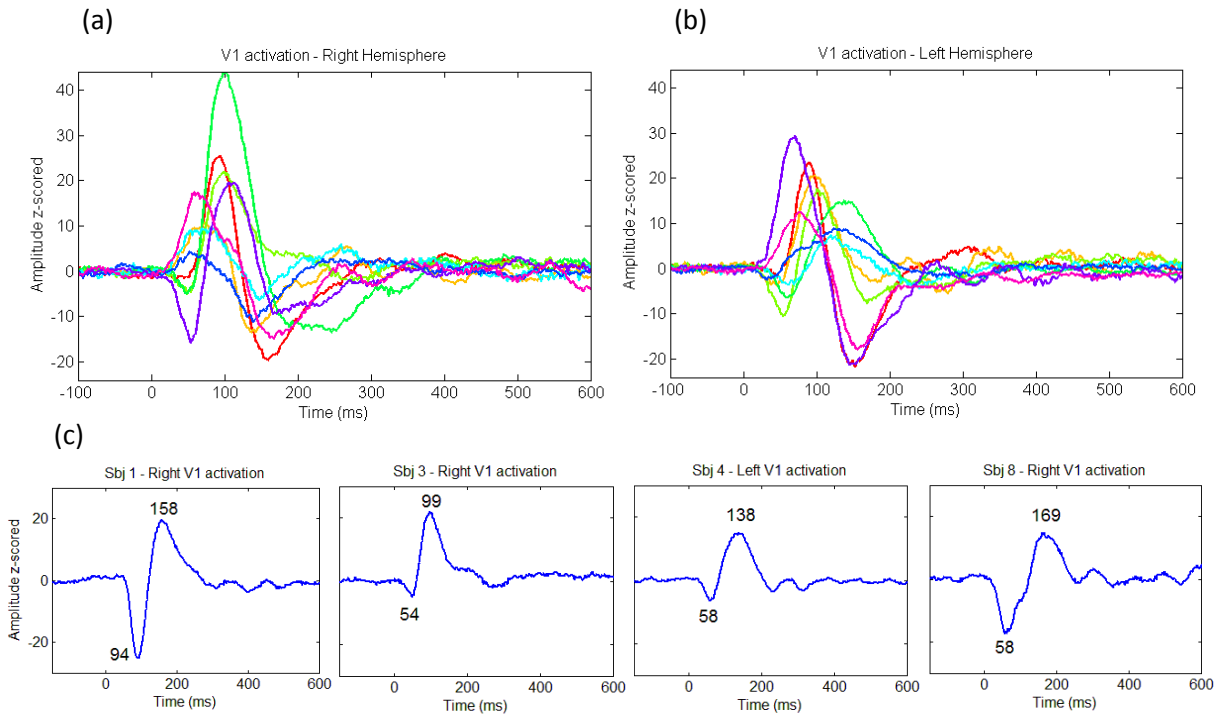

Fig. 3-3: Time-courses of the VER obtained from area V1. (a) and (b) show the superposition of VEPs for individual subjects in the right and left V1 respectively. The visual activation is averaged over all conditions of natural image stimulation (34ms presentation). (c) shows the VEPs for 4 different subjects along with the timings of the first two clear peaks.

The difference between the VEPs from the two hemispheres of one subject is presented in Fig. 3-4. This figure illustrates is based on the same data as Fig. 3-3, , but portrays the data in a different manner, i.e. the right (panel (a)) and left (panel (b)) hemisphere data of a single subject. In the left hemisphere (panel (b)), all response peaks correspond to the typical C1, M100, and M170 peaks. However, this expected pattern is not observed in the right hemisphere's response (panel (a)).

Due to these apparent differences, hemisphere responses were further examined using standard visual stimuli of typical VEP studies, such as the checkerboard stimuli. It is important to note that only preliminary analysis of these datasets was performed since further analysis was deemed

unnecessary for the purpose of the current study. The visual evoked responses to the checkerboard stimuli (red curves) overlaid on the responses to natural images (blue curves) is depicted by Fig. 3-5. Panels (a) and (b) represent the data from two participants, including right and left hemisphere responses to 'short' (34ms) and 'long' (400ms) stimulations. As shown by the figure, the shape and peak times of the VERs were nearly identical for checkerboard and natural image stimuli.

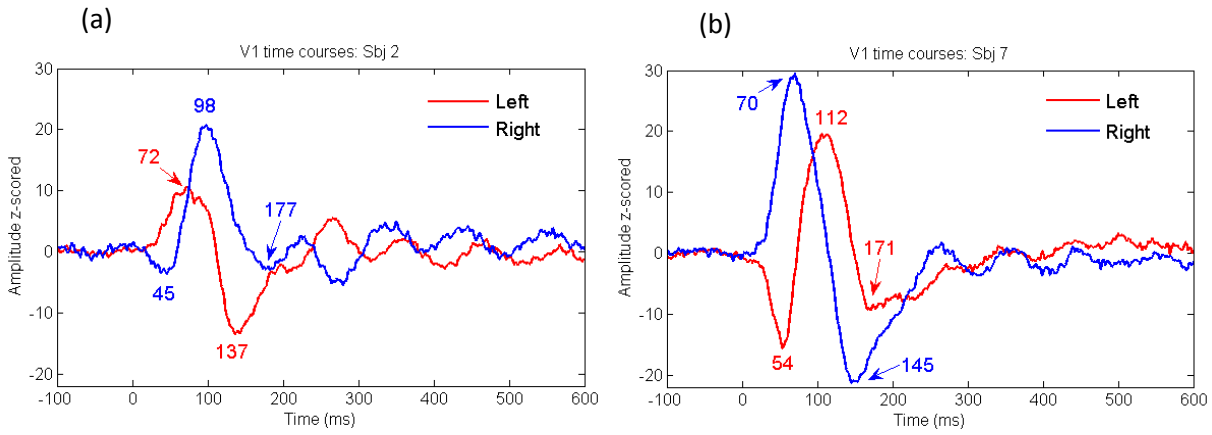

Fig. 3-4: VER localized in area V1 of the right (red) and left (blue) hemispheres from two subjects. It is clear that even in one subject and between two hemispheres there may be significant differences in timings of the peaks and shape of the MEG-based VER.

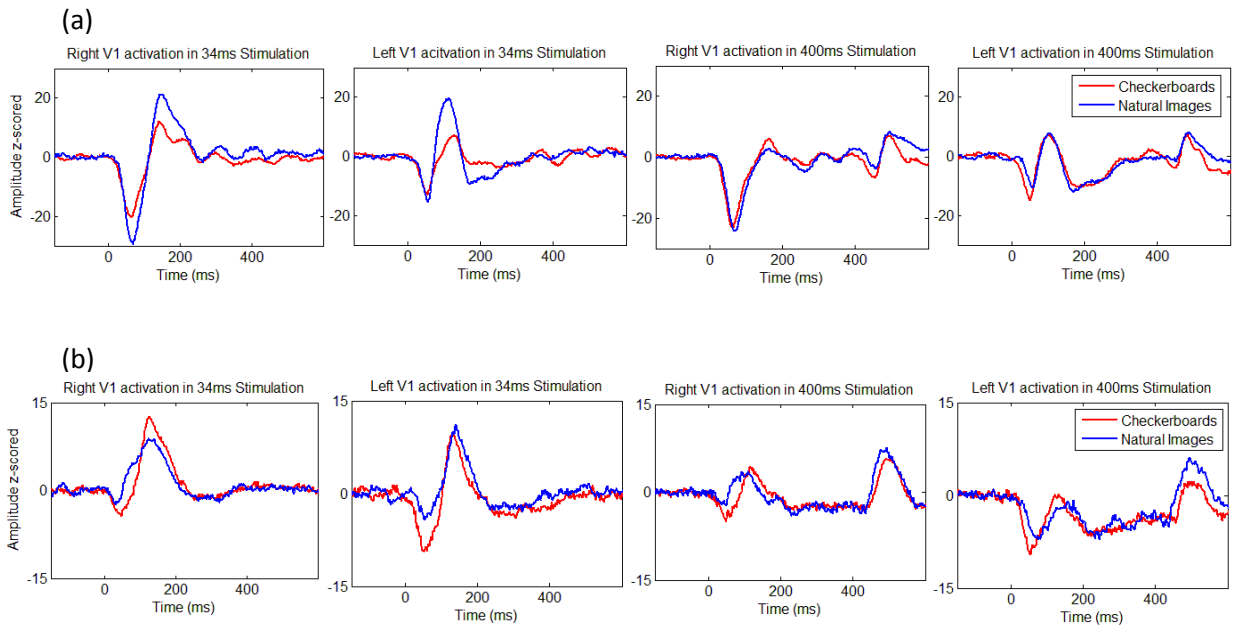

Fig. 3-5: VER localized in area V1 in response to Checkerboard stimuli (red curves) and natural Images (blue curves) stimuli. Panels (a) and (b) present the results from two subjects. Each panel includes the response in left and right hemispheres and for long and short duration of stimulation. Without considering the amplitude of the signals, the shapes and the peak timings of both VERs are nearly similar in each hemisphere and each subject.

3.2. MEG results

3.2.1. Eye-tracker analysis results

Three samples of eye-fixation performance, as measured by the eye-tracker, are represented by Fig. 3-6. The top row depicts the raw sample points while the bottom row depicts the corresponding heat maps of samples distribution. Two panels ((a) and (b)) portrayed good eye fixation with few saccadic movements in panel (b), whereas panel (c) shows constant eye-movements around the fixation point. It is important to note that panel (c) represents a rare case within the dataset of this study (one subject and only one run). Aside from this subject, almost all subjects performed acceptable fixation throughout the entirety of this experiment.

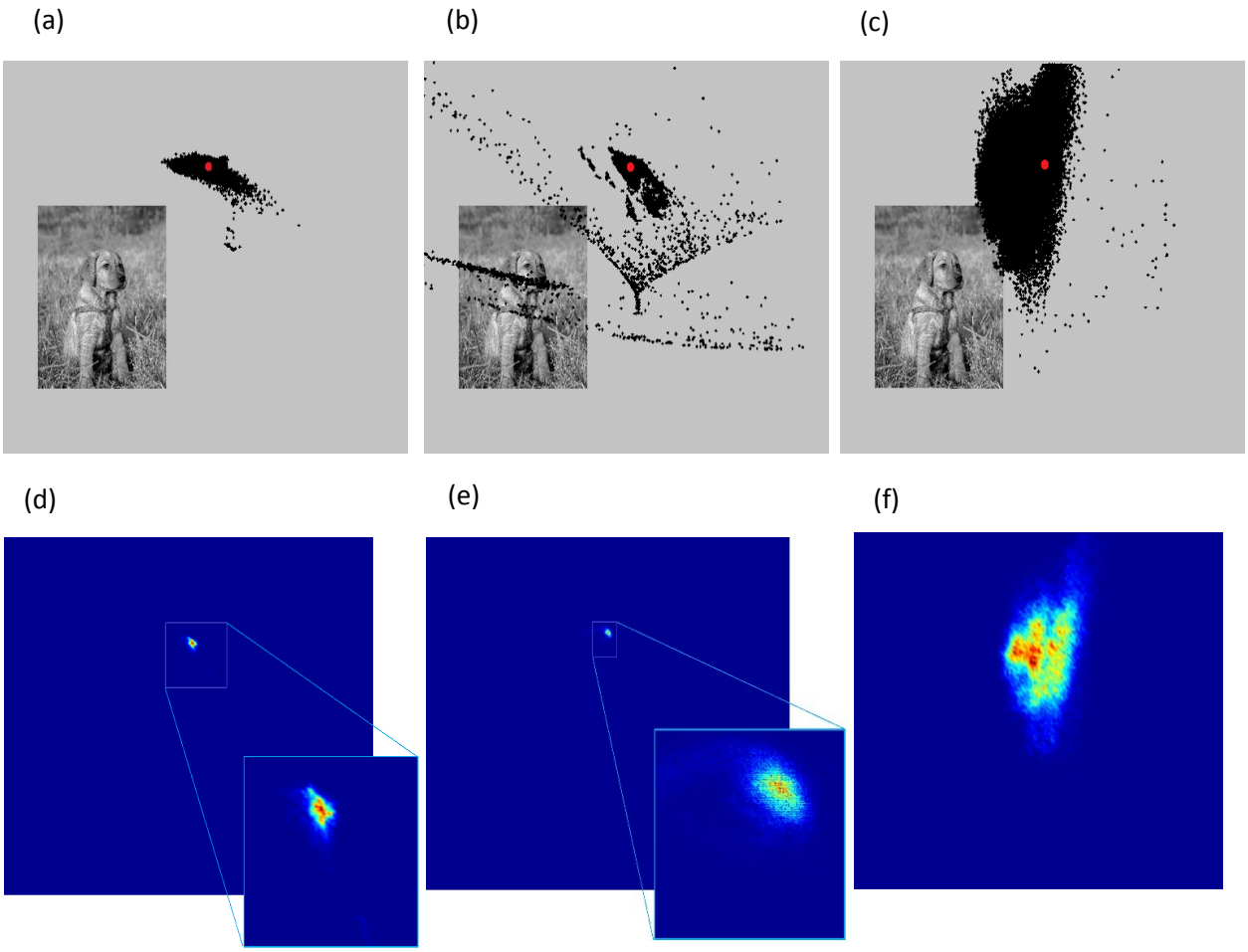

Fig. 3-6: Three examples of eye-fixation performance. The top panels show the eye-tracker sample points for one run (11 minutes) overlaid on a schematic screen. The panels at the bottom show the heat-maps of the distributions relative to the visual stimuli (same data as those used for the top panels). In (d) and (e) parts of the maps were zoomed-in on the right corner of the figures. (a) and (d) show an example of very good fixation with no large saccades. In example (b) and (e), the subject performed well as shown in the heat-map. However, he made few saccadic eye-movements of large amplitudes. In contrast, (c) and (f) show data with constant eye-movements around the fixation point.

Fig. 3-7 shows the results of the saccade detection method used on some extracted epochs. Red shaded areas correspond to macro-saccades (larger than 1 degree), which were detected using the ClusterFix package. As shown by this figure, this method successfully captured different forms and durations of saccadic movements.

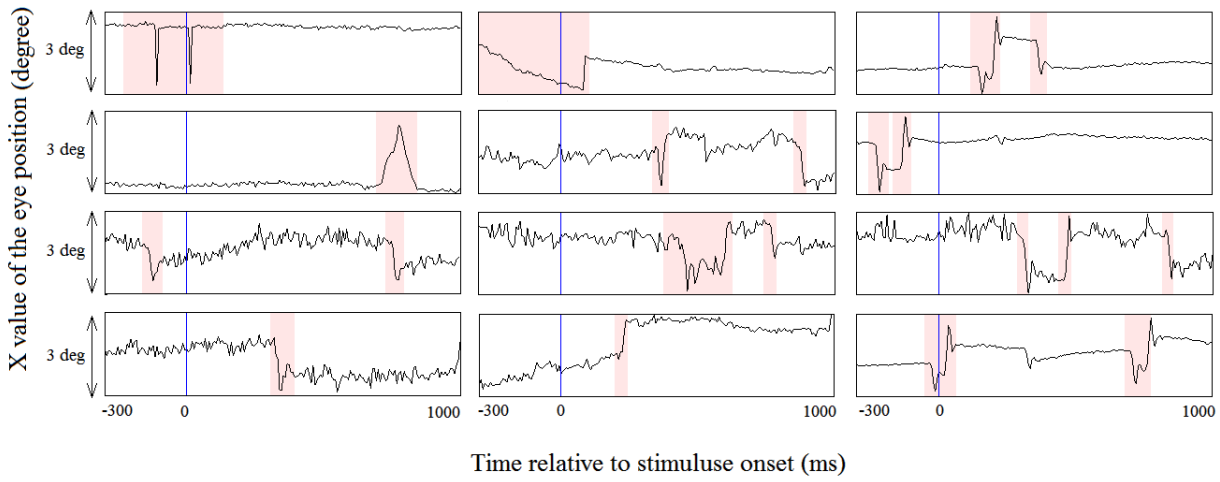

Fig. 3-7: Samples of saccade detection results on epochs of eye-tracker data. The epochs were gathered from different runs of different subjects. Blue vertical line corresponds to the on-set of the visual stimulation and the red shaded blocks indicates the detected duration of macro-saccadic (larger than 1 degree) eye movements.

The distribution of macro-saccadic eye movements throughout the experimental trials is shown in Fig. 3-8. The percentage of trials containing macro-saccades at each specific time was computed for each subject. These time courses were then averaged over seven subjects and illustrate by different colors corresponding to different stimulus conditions. A one-factor within-subjects ANOVA was performed on each time point of this data and failed to reveal any significant difference between saccadic eye movements in different stimulus condition ($p>0.8$). The possibility of biased eye movements within individual subjects was investigated using a one-factor within-subjects ANOVA on the standard deviation of eye position at each time point. This analysis revealed that none of the subjects showed significant difference across stimulus conditions with respect to eye movements ($p>0.05$).

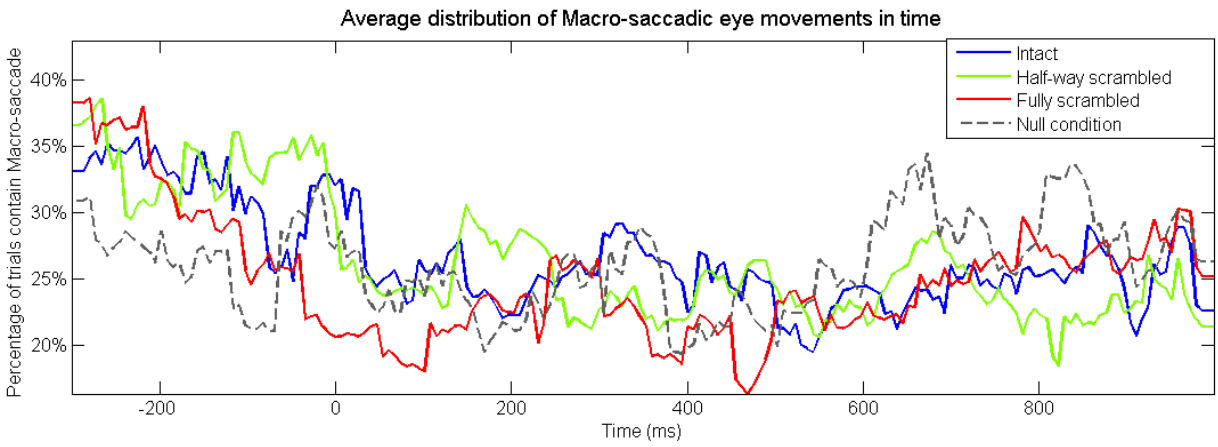

Fig. 3-8: Average distribution of Macro-saccadic eye movements in time for different conditions of the stimulus (400ms presentation). Time zero indicates the stimulus onset and the eye movements were concatenated over the period of -300ms to 1000ms post-stimulation. At each time point, the percentage of trials which contain macro-saccade at that time point was computed for each condition of the stimulus and each subject. The graph shows the average of these values over 7 subjects (one subject was excluded because of eye-tracker recording problem). The dashed gray curve indicates the macro-saccadic eye movements during the Null condition, i.e. in the absence of visual stimulation. The result of one factor within subject ANOVA performed on each time point, failed to reveal any significant difference between the eye movements in different stimulus conditions (p>0.8).

3.2.2. Subjects' response analysis results

A one-factor within-subjects ANOVA was performed on the subject's response and failed to reveal any significant difference between subjects' performance in different stimulus conditions ($p>0.4$). Fig. 3-9 shows the average percentage of correct response in the fixation task for the different conditions.

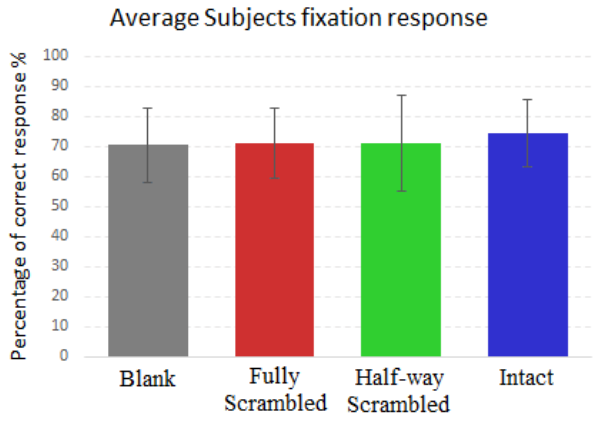

Fig. 3-9: The average of the subjects' performance in the fixation task within different stimulus conditions. The error bars indicate the standard deviation between subjects. No significant difference has been observed between conditions.

In addition, histograms of response delays for one subject are represented by Fig. 3-10. Different panels correspond to changes occurring during different stimulus conditions. The red vertical line indicates the threshold of acceptable response delay, which was set as 0.5s. The Kolmogorov-Smirnov test between pairs of histograms was performed for each subject. In two subjects, the test revealed a significant difference between blank and image-presenting conditions ($P<0.05$). However, no significant differences were found between the three conditions of image presentation ($P>0.1$).

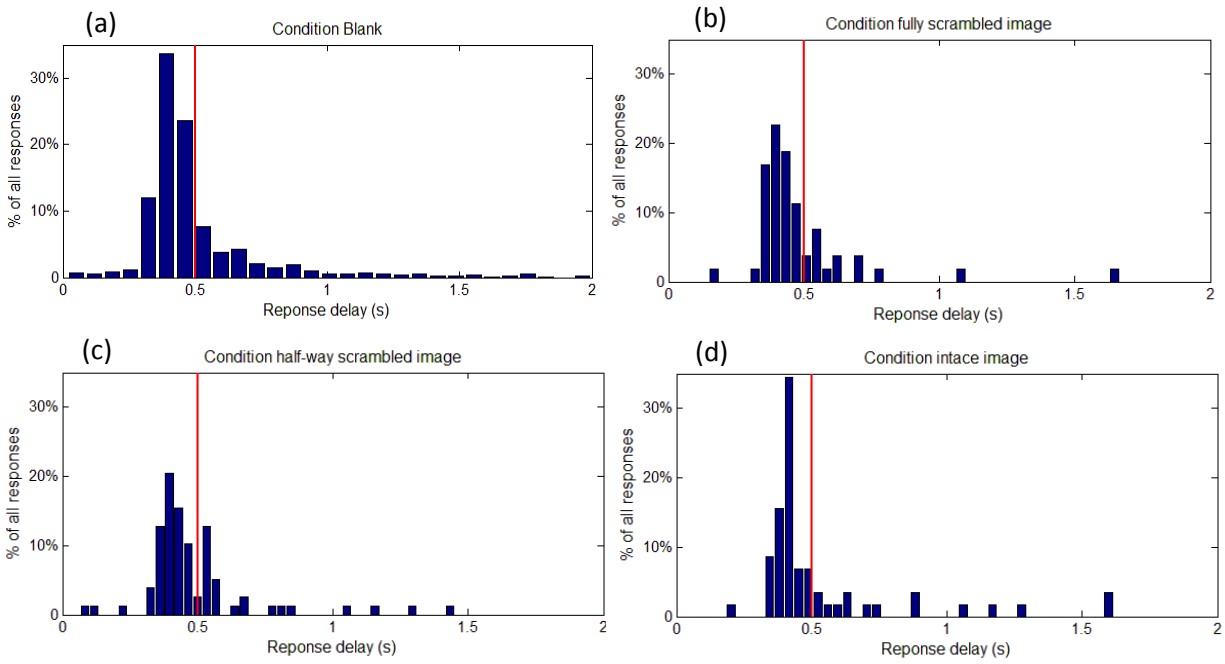

Fig. 3-10: Response delay histograms from one subject, in four different conditions of the stimulus. The panels present the responses to the changes taking place during (a) blank, (b) fully scrambled image (c) half-way scrambled image, and (d) the intact image condition. The red vertical line indicates the delay threshold for unacceptable responses (0.5s).

3.2.3. Results of the MEG sensor space analysis

As previously described, the RMS of 40 sensors with the highest peak-to-peak amplitudes was computed for each subject. Fig. 3-11 and Fig. 3-12 show the average of this RMS signal computed over subjects for right (a) and left (b) hemisphere stimulation, as well as 'short' (Fig. 3-11) and 'long' (Fig. 3-12) stimulus durations. The results of post-hoc pairwise comparison tests are shown with white, black, and gray bars next to the horizontal axes of the plots. The early peak of activation is clearer in Fig. 3-11(a) and is significantly different in the intact condition compared to the other two conditions (halfway and fully scrambled). Another time period with significant effect of image scrambling in the short stimulation response is seen

from 100ms to 200ms post-stimulation in both the right and left hemispheres (Fig. 3-11(a) and (b)).

In addition, Panels (c) and (d) in Fig. 3-11 and Fig. 3-12 summarize the statistically significant differences within specified time epochs. In the 'short' stimulation condition (Fig. 3-11), the smaller averaged RMS value in the intact condition compared to scrambled conditions is highly significant in 140-200ms period in both hemispheres ($p<0.003$). A similar difference was observed in the early epoch, 30-70ms, for the right hemisphere as well as between the intact and fully scrambled conditions in the left hemisphere ($p<0.05$). A significantly larger RMS value in the intact condition, compared to fully scrambled condition, within the 80-120ms time epoch can be observed only in the right hemisphere ($p<0.02$).

The 'long' stimulation results consist of short time periods with significant differences between the fully scrambled and halfway scrambled conditions, as depicted by the white lines in Fig. 3-12. The only difference between the right and left hemisphere responses in the 'long' stimulus presentation is the significantly larger averaged RMS value in the halfway scrambled condition compared to the intact image condition for the 30-70ms time period ($p<0.02$). The two hemispheres show no significant differences between conditions for the 80-120ms time period. However, during the 250-400ms post-stimulation time period and the off-response (450-510ms) time period, the RMS values are significantly larger for the intact image condition compared to the scrambled counterparts ($p<0.0003$ in left hemisphere and $p<0.03$ in right hemisphere).

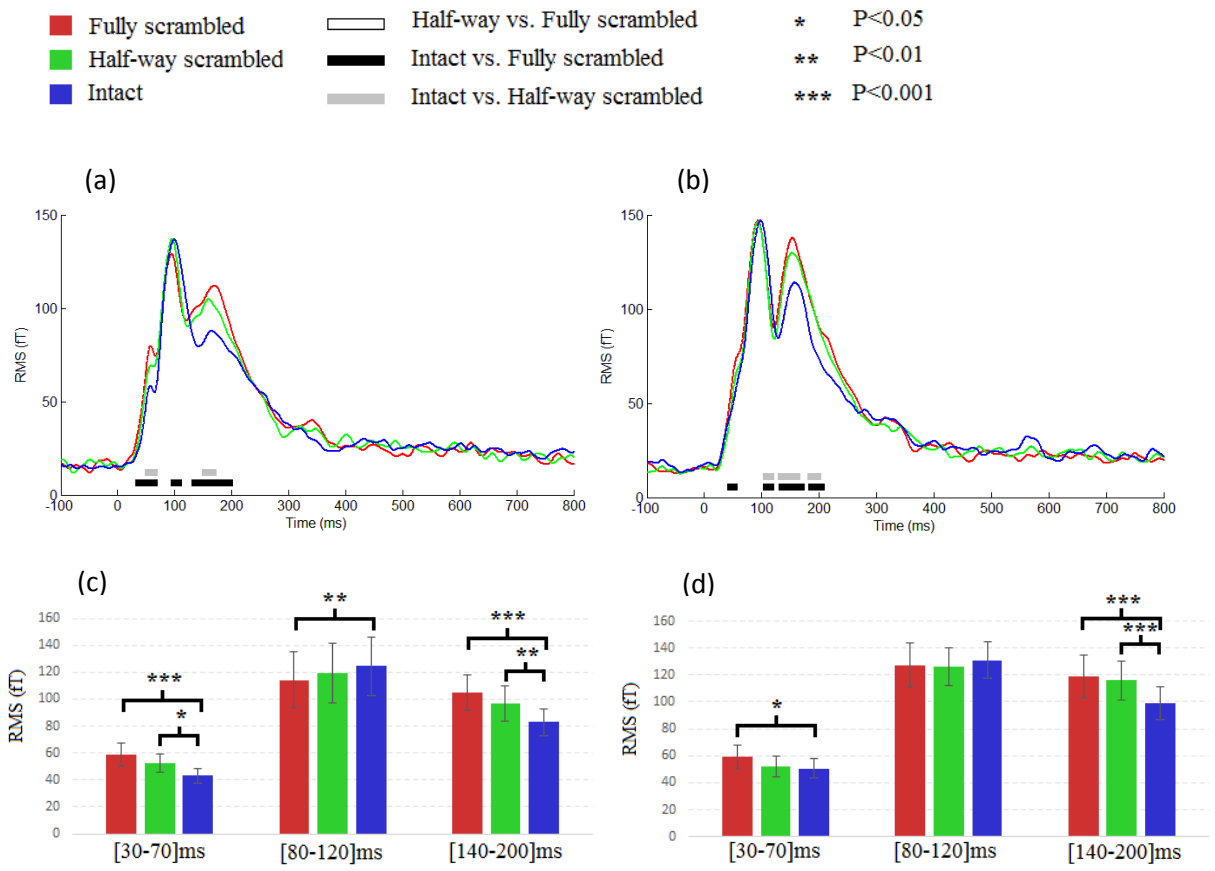

Fig. 3-11: Comparing Channels RMS signal between different conditions of the short (34ms) stimulus. The RMS was computed over 40 channels with the highest peak-to-peak amplitude in each subject. Panel (a) and (b) correspond to the activation of the right and left hemispheres respectively. The white, black and gray bars near the horizontal axes indicate the periods of significant pairwise differences between condition. Panel (c) and (d) show the averaged RMS value within the specified time epochs and are obtained from the right and left hemisphere responses, respectively. Error bars correspond to the standard error of the mean over the 8 subjects. Significant pairwise differences are also indicated on the bar plots.

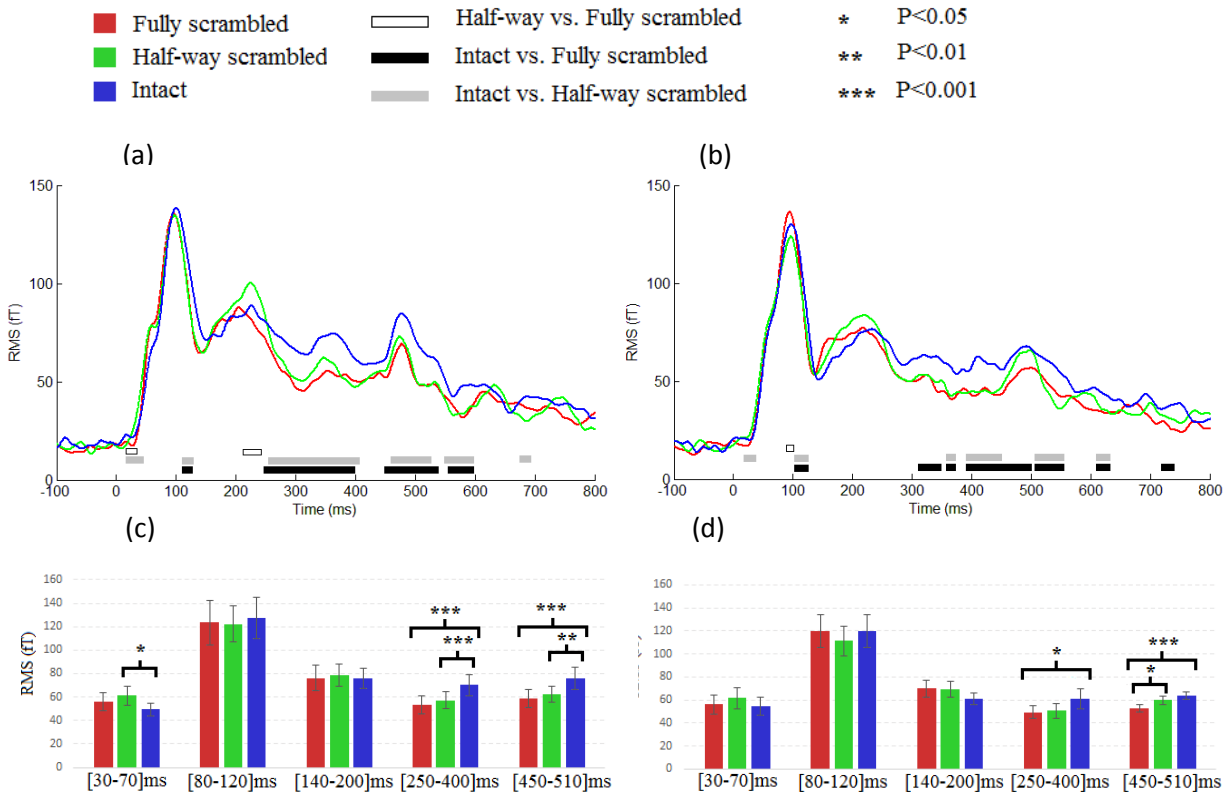

Fig. 3-12: Comparing Channels RMS signal between different conditions of the long (400ms) stimulus. The RMS was computed over 40 channels with the highest peak-to-peak amplitude in each subject. Panels (a) and (b) present the activation of the right and left hemispheres, respectively. The white, black and gray bars near the horizontal axes indicate the periods of significant pairwise differences between conditions. Panel (c) and (d) show the averaged RMS value within specified time epochs. They were obtained from the right and left hemisphere responses, respectively. Error bars correspond to the standard error of mean computed over the 8 subjects. Significant pairwise differences are indicated on the bar plots too.

3.2.4. MEG source space results

For the source space analysis, in which a two-factor within-subjects ANOVA test was performed, the responses of both hemispheres were regarded as a within-subjects factor ('hemisphere'). In doing so, the power of the test was increased. The results of the V1 time course analysis are depicted by Fig. 3-13. . Panel (a) shows the average VEP response of area V1 for different stimulus conditions. Activation of V1 in response to visual stimulation is represented by dashed and solid curves for 'short' and 'long' conditions, respectively. Gray-scale

lines in the top and bottom portions of the graph specify the significant pairwise differences between the condition for the 'short' and 'long' presentation, respectively. Fig. 3-13(b) and (c) demonstrate the averaged z-scored amplitudes within the specified time epochs. No significant difference between scrambling levels can be seen in the early response (30-70ms). However, the averaged VEP value within 140-200ms is significantly different between the intact and scrambled conditions ($p<0.002$). The signal response associated with the scrambled condition remains higher than the response to the intact image condition, irrespective of the sign. This trend prevails in the following time epochs (250-510ms) for the 'long' stimulus condition (Fig. 3-13 (c)) ($p<0.002$).

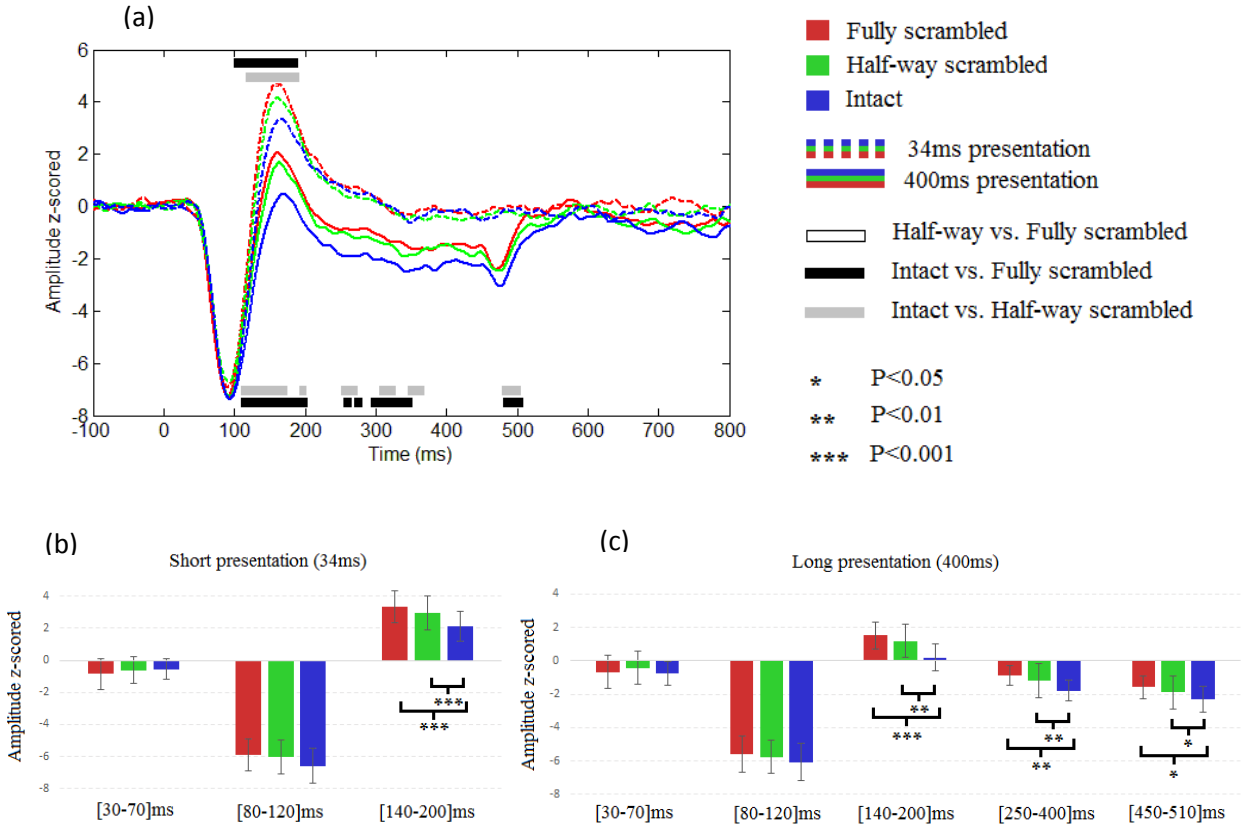

Fig. 3-13: Comparing V1 evoked potential between different conditions of the short (dashed lines) and long (solid lines) stimuli. Panel (a) shows the average VEPs evoked by the different conditions, computed over subjects and hemispheres. The black and

gray bars near the horizontal axes indicate the periods of significant pairwise differences between conditions. The bars in the bottom of the figure correspond to the long stimulations (solid lines) and the ones on top are related to the short stimulations (dashed lines). Panel (b) and (c) show the averaged amplitude values within specified time epochs. Error bars reflect the standard error of mean for 16 samples (8 subject and 2 hemispheres) Significant pairwise differences are also indicated on the bar plots.

No significant differences are observed in the early response of V1 activation Fig. 3-13. More specifically, the C1 component cannot be captured in the average VEP signal, as depicted by Fig. 3-13(a). The absence of the C1 component could potentially be due to variability among subjects in the exact timing of this component. In order to compensate for this variability, the responses were aligned according to their peaks within the 50 to 60ms post-stimulation time period. The average of these aligned responses enables the identification of the early peak at approximately 55ms for both 'short' and 'long' stimulation triggered responses (Fig. 3-14 (a) and (b)). Although no significant differences are observed between the two levels of phase scrambling, the amplitude of the intact condition is significantly smaller relative to the scrambled conditions ($p<0.002$).

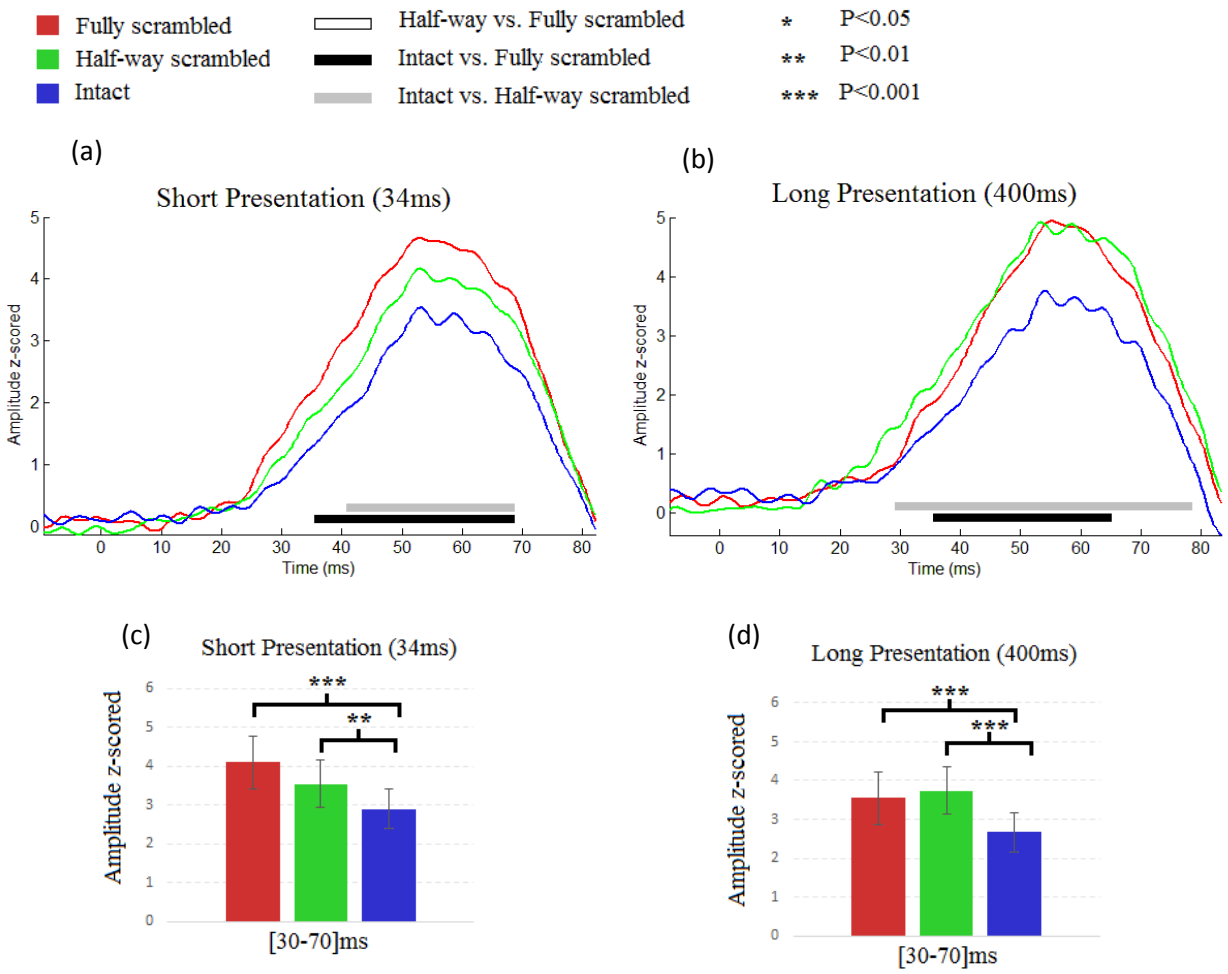

Fig. 3-14: Comparing the early peak of V1 evoked potential between different conditions of the short ((a), (c)) and long ((b), (d)) stimuli. Panels (a) and (b) show the average of the VEPs among subjects and hemispheres after aligning the first peak of the VEP of different samples. The black and gray bars near the horizontal axes indicate the periods of significant pairwise differences between conditions. Panels (c) and (d) show the averaged amplitude values within the time period of 30 to 70ms. Error bars reflect the standard error of mean over 16 samples (8 subjects x 2 hemispheres). Significant pairwise differences are also indicated on the bar plots.

The results of the time-frequency analysis in the source space are shown in Fig. 3-15. The time-frequency responses to the intact and scrambled conditions were compared, and the difference between these responses was used to derive the top panel representation in Fig. 3-15. Colors closer to the red end of the spectrum indicate regions in which the power of a frequency band was higher for the scrambled image condition. Meanwhile, colors closer to the blue end of the

spectrum represent regions in which the power of a frequency band was higher for the intact image condition. The horizontal line indicates the onset of the stimulus at point 0. Panels (c) and (d) correspond to the same time-frequency maps portrayed in (a) and (b), respectively. However, panels (c) and (d) only include spots that show significant difference between the intact and fully scrambled conditions ($p<0.05$ corrected). Gray areas represent time-frequency points with a non-significant difference ($p>0.05$ corrected) between these two conditions.

As shown by Fig. 3-15(b), the scrambled condition induces relatively stronger gamma power compared to the intact condition (from about 40Hz up to 110Hz) with respect to the M170 (100-200ms post-stimulation) component in response to the 'long' presentation. On the other hand, the intact condition shows stronger power in the beta band at this time period (b-II). Higher power of high frequencies, 60-80Hz, in the scrambled condition seems to continue until the end of image presentation (b-III). Though, during the off-response time point, a significantly stronger 50-60Hz power emerges in the intact condition (b-IV). This difference can also be seen for responses to the 'short' presentation (a-IV). Further, the beta power for the intact condition following the offset of the stimulus is significantly higher for both 'short' and 'long' presentation conditions (a-V and b-V).

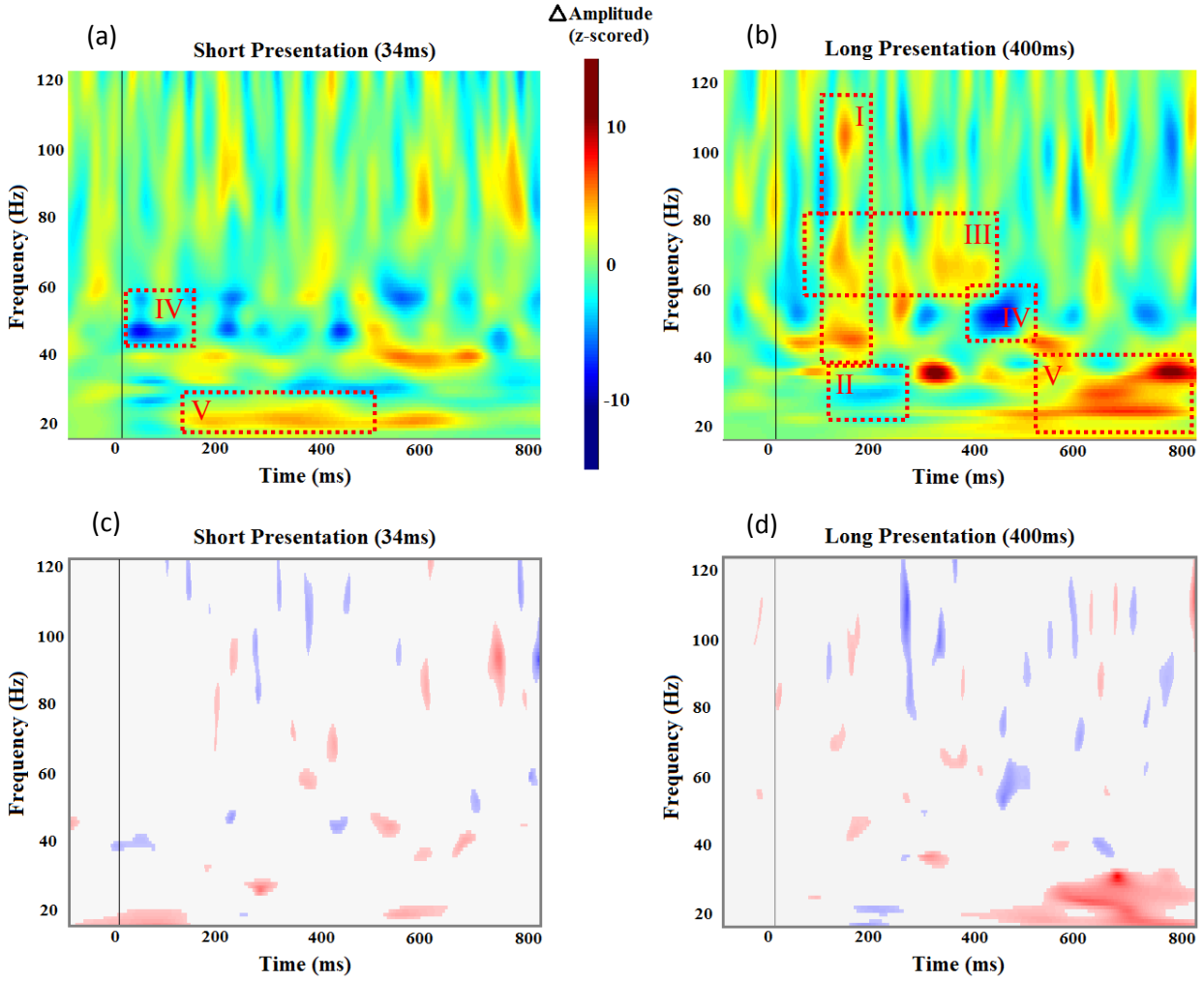

Fig. 3-15: Time-frequency decomposition of the activity localized in V1 for short (a) and long (b) presentations. The z-scored power of frequencies less that 16Hz cannot be computed due to the edge effect in time-frequency decomposition and are not presented in this figure. Panels (a) and (b) show the difference between the time-frequency responses to the fully scrambled condition and the intact condition. Yellow and red colors indicate the spots where the scrambled image condition had higher power, whereas the blueish areas indicate higher power in the intact condition. Black vertical lines indicate the onset of the stimulus. The time-frequency maps were z-scored relative to the 200ms before the onset. Panels (c) and (d) show the result of a paired t-test between the fully scrambled and intact condition using 8 subjects data. The maps were thresholded at p –value equals to 0.05. (Red rectangles and the Roman numbers are for the illustration in the text.)

4. Discussion

This study examines the modulation of activity in the primary visual cortex (V1) in response to changes in the higher order statistics of natural images. The main finding of this study confirms the initial hypothesis that the population activity in V1 is modulated via the manipulation of the phase content of visual input. This modulation is discussed in the following section, which also refers to three other specific findings: the early effect of phase scrambling on the Evoked Response Potential, the change of the M170 component amplitude, and the difference in the induced power of specific frequency bands.

4.1. Early ERP modulation

Image phase scrambling significantly modulated brain activity as early as 50ms post-stimulation. Scrambling the phase content of visual input is shown to increase in the magnetic field power of the early response (30-70ms post-stimulation) (Fig. 3-11, Fig. 3-12). These effects are also observed in the visual evoked response of area V1 after performing source localization (Fig. 3-14). However, V2 activity may potentially contribute to the time-courses obtained from the ROI in area V1 due to imperfect spatial resolution and deficient specificity of the MEG source reconstruction method.

The early component of the VEP signal (known as C1 in EEG studies) has been subject to several studies using visual stimuli. The C1 component was once considered to predominantly reflect the activity of V1 (Di Russo et al., 2005; Kelly et al., 2013). However, several studies have pointed out the possibility that V2 and extrastriate areas contribute to the early peak of the

VEP (Ales et al., 2010; Foxe and Simpson, 2002). Using high density electrical mapping in human, Foxe and Simpson (2002) showed that neural activity propagates rapidly during the C1 interval, whereby the dorsal stream response preceding that of the ventral stream. According to this study, activity in the frontal cortex can be observed as early as 30ms after the onset of the C1 response, while the extrastriate areas can be activated even earlier. Thus, only the very early phase of C1 (the first 10-15ms) can exclusively reflect V1 activation.

Considering the findings of Foxe and Simpson (2002), our results certainly indicate the potential contributions of extrastriate areas to the early component of the VEP signal. Each panel in Fig. 4-1 depicts the source map average within the 30-70ms time period and within 8 subjects, which was derived following the registration of each source map onto the standard anatomy (Colin 27). According to this figure, the source of the early visual response (after averaging all image conditions) is localized in areas V1 and V2 (Fig. 4-1). Thus, it is possible to interpret the modulatory effect of the early VEP based on the feedback projections from extrastriate areas to the primary visual cortex.

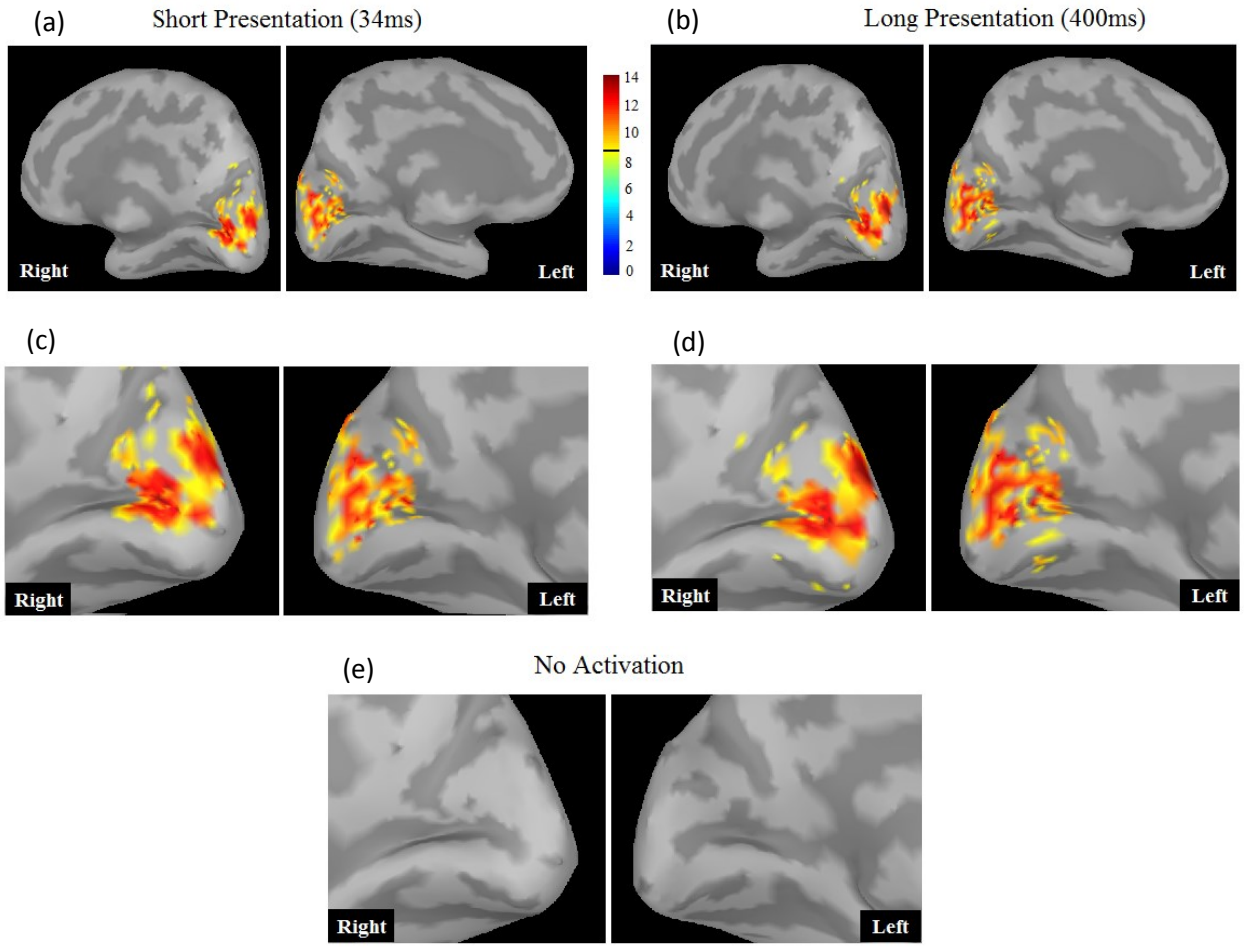

Fig. 4-1: localization of the visual activation (the average response over all image conditions) within 30-70ms post-stimulus. Panel (a)-(d) show the average of 8 subject's source maps projected on the standard anatomy (Colin27) and also averaged in time from 30-70ms post-stimulus. The color bar shows the activation values z-scored with respect to baseline. The maps show sources with maximum response higher than 60% of the maximum response over space (black line in the color bar). (a) Represents the activation for short presentation (34ms) and (b) corresponds to the activation for long presentation (400ms). (c) and (d) are the magnified version of (a) and (b) respectively. Panel (e) represent the anatomical view of the visual cortex without any active source.

This early modulation of neural activity could be explained by the sparse coding theory of stimulus representation in primate area V1. Scrambling the phase of the natural images eliminates spatial structures that are important factors for sparse neural code. Natural images elicit weaker activity because fewer neurons need to be simultaneously active to code a sparser

input. In other words, V1 activity is not only sensitive to the second-order statistics (RMS contrast) of the input, but also to the spatial phase and the fourth-order statistics (kurtosis).

In contrast to Rieger et al.'s study, the results of this study show an increase in V1 activity as sparseness (kurtosis) decreases (Fig. 2-4). These two studies are mainly differentiated by their approach of analyzing the acquired data. In Rieger et al.'s study, the research team compared the signals between intact and phase scrambled conditions for every single MEG channel. In contrast, in this study, the RMS of a subset of 40 MEG channels was computed and then compared between different stimulus conditions. This method of data analysis increases the SNR and detects more powerfully the small, yet significant, effect of phase scrambling on early VEP signals.

Moreover, in the source space analysis, Rieger et al. investigated the time courses of ROIs in area V1 from 50-100ms post-stimulation. Similarly, the results of this study did not report any modulatory effect at this time period in the preliminary analysis (Fig. 3-13). However, after taking into account the individual variability in the shape of the VEP response by aligning the peak of the early activation in V1 (Fig. 3-14) – a step lacking in Rieger et al.'s analysis – a significant difference is uncovered.

Further, the interpretation of these results with respect to the sparse coding of natural images coincides with the findings from a recent study in mice (Froudarakis et al., 2014). In this study, the authors employed natural and phase-scrambled movies as visual stimuli. They reported an inverse correlation between the number of neurons that code the information and the spatial correlation in natural scenes. In other words, as spatial correlation increased, the number of neurons coding the information actually decreased. Several other studies have also reported

sparser neuronal activity in response to natural versus noise inputs (Dumoulin and Hess, 2006; Rainer et al., 2001).

The sparse response to natural images can be explained by the sparseness of natural images themselves, which can be assessed according to the distribution of local contrast in the image. Fig. 4-2 shows the outcome of first-order directional Gaussian derivative filter applied to the images. The method explained in Scholte et al. (2009) was used in this study to extract filtered images and plot the local contrast histograms (Fig. 4-2 (b)) (Scholte et al., 2009). Similar to the results reported by Scholte et al. (2009) natural images show a distribution that peaks at low edge strength while the strong edges are distributed sparsely throughout the image. In contrast, in the phase-scrambled versions of the image, the contrast is more evenly distributed and results in a wider distribution of edges.

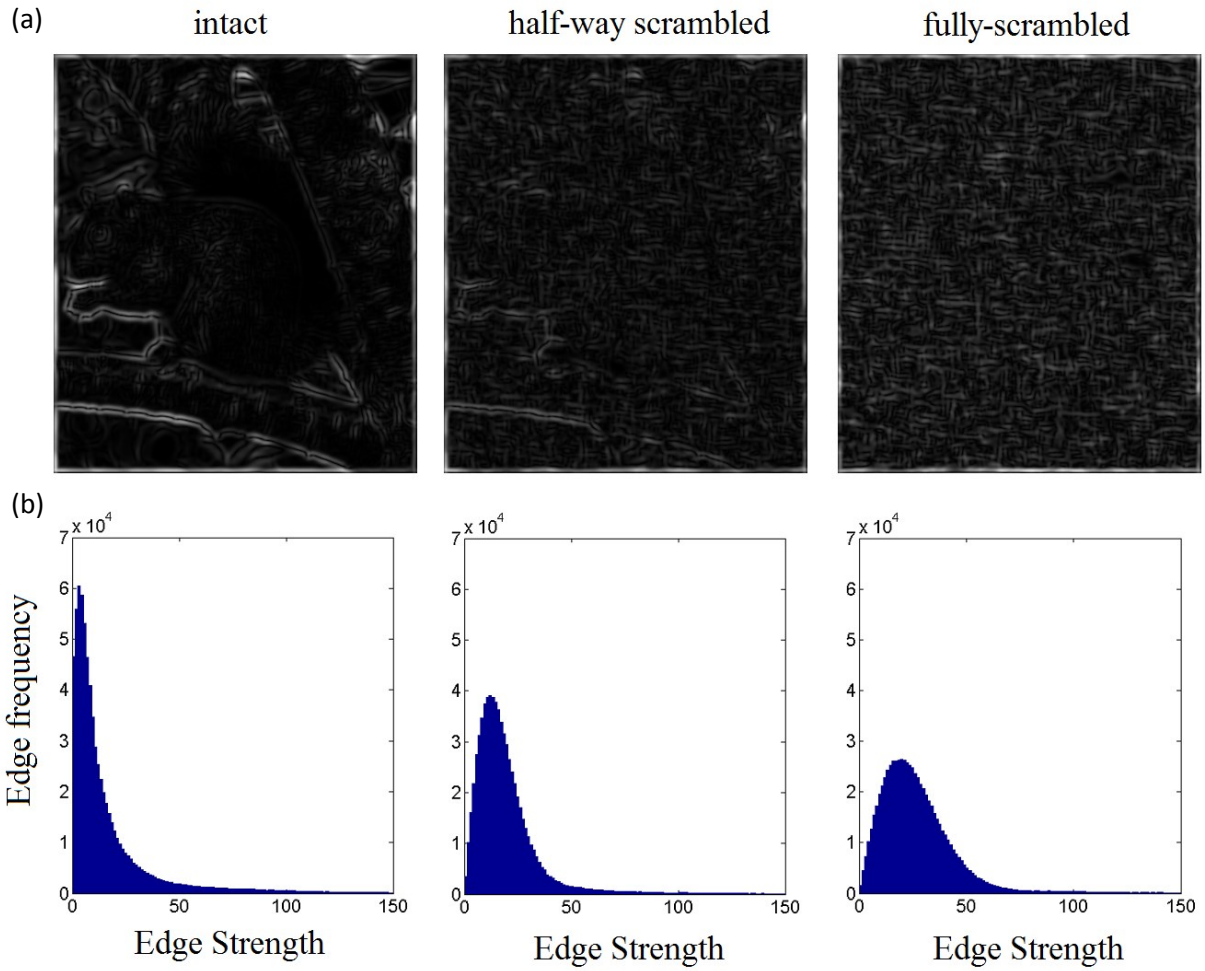

Fig. 4-2: The effect of phase scrambling on the distribution of image local contrast. (a) The result of applying first-order directional Gaussian derivative filter to the images. The response of the filter is maximal in the location of edges. (b) Distribution of images in (a). Natural images have edge distribution which peaks edges of low strength, whereas in phase scrambled images the contrast is distributed more smoothly.

Other recent studies have also reported the modulation in the early VEP component. In Poghosyan and Ioannides (2008), the researchers showed that attention enhances the initial response in primary visual and auditory cortices at 55-90ms and 30-50ms post-stimulation, respectively (Poghosyan and Ioannides, 2008). Further, the use of a visual memory task in an MEG study, Chaumon et al. (2008) show that memory can influence the early response (before

100ms) to visual stimuli (Chaumon et al., 2008). In addition, in Rauss et al.'s review (2011), the evidence of top-down modulatory effects during visual processing is described in detail. These authors hypothesize that earlier studies fail to detect this early modulatory effect due to the fact that its identification requires a specific experimental setup, including stimulus design and factors to account for variability between subjects (Rauss et al., 2011).

The early visual response in area V1 is known to vary between individuals in its shape and area, resulting in different C1 components across subjects (Kelly et al., 2008). Based on Kelly et al.'s study, only a small number of subjects tend to exhibit robust C1 from stimulation in a single location of visual field. In turn, they propose that a probing session be preformed prior to data acquisition, which would enable the optimization of C1 recognition for each subject. In this session, researchers would be able to identify both the region of the visual field that elicits the greatest response and the precise EEG sensors that most strongly detect C1 activity. Based on this method, they could find an enhancement in the C1 component due to the effect of attention as early as the onset of the C1 component (57ms) (Kelly et al., 2008).

In this study, the stimulus presentation was constrained to the lower part of the visual field. In doing so, only the upper portion of the calcarine sulcus is activated, resulting in more prominent C1 components. If a visual stimulus were presented to the entire visual field, both the upper and lower portions of the calcarine sulcus would be activated. This activation would result in inversely oriented electrical dipoles that could partially or fully mask the opposing signal and reduce the overall signal recorded at the sensor level. In addition, the first component of the VEPS in the source level was aligned prior to their averaging and use in statistical tests. By adding this step in the data analysis, individual variability is accounted for.

4.2. Modulation of the M170 component

The modulatory effect on visual activation can also be seen at the M170 component, ranging from 140-200ms post-stimulation. Similar to the early visual activation, the M170 component is higher for phase-scrambled conditions compared to the intact image condition. By this time period, visual information has reached higher-order areas and undergone high-level processing stages. In addition, according to Fig. 3-13(a) the peak in the M170 component is higher for 'short' stimulations relative to 'long' stimulations. The underlying mechanism for this phenomenon is believed to involve surround suppression, which could be stronger when a stimulus is presented continuously for a longer time.

Similar to this interpretation, the presence of elongated edges in natural images could potentially weaken the M170 component. According to previous studies, natural images are coded efficiently by a lower number of active neurons compared to their unnatural counterparts. Moreover, edges activate neuronal populations that show preference to their orientation and spatial frequency. In natural images, elongated edges with few abrupt changes in orientation are common. Therefore, in response to these images, the population of neurons that exhibit preference to this edge orientation may have receptive fields that are close to one another. In turn, since it would be redundant for multiple neurons to code for nearly identical information about local sub-portions of the natural image, it is hypothesized that these neurons suppress each other. However, in phase-scrambled images, the spatial coherence of the image is removed, rendering nearby populations of neurons to represent different edge information. Thus, surround suppression would eventually be weaker for phase-scrambled images compared to their natural counterparts.

Aside from the potential involvement of lateral suppression, the modulation of the M170 component can be interpreted based on the concept of Predictive Coding. According to this model, top-down expectations suppress the response to the expected input. Since visual information processing has evolved according to natural inputs and their higher-order statistics, prediction errors introduced in lower visual areas are hypothesized to be higher under phase-scrambled presentation conditions compared to natural image presentation. The observed significantly higher amplitude of the M170 component in response to phase-scrambled images relative to intact images coincides with the aforementioned hypothesis. This modulation, which was observed in area V1, could be representative of the emergence of prediction error in this area. The same interpretation was adopted by (Todorovic and de Lange, 2012) for the neural response to an unexpected auditory tone. The research team related this increase in the evoked response potential to higher prediction error in the intermediate stages of auditory processing, 100-200ms post-stimulation.

Our findings are consistent with those reported in Huberle et al.'s MEG study, in which images of objects, both intact and tile-scrambled, were presented to subjects. The research team localized the neural activation by fitting four dipoles in the source space. Their results show a higher response to scrambled images than to intact images for 70-170ms post-stimulation, which was localized in early visual cortex. Moreover, they reported higher responses to intact images in later stages, 200-270ms post-stimulation. The latter responses were observed in the majority of occipital and temporal sensors as well as several frontal and central sensors (Huberle and Lutzenberger, 2013). However, it is important to note that the stimulus of this study is more controllable than the one used in Huberle and Lutzenberger (2013). As mentioned previously, the

phase-scrambling approach maintains the global contrast while altering the higher-order statistics of the image, whereas the tile-scrambling process (employed by Huberle et al.) introduces high contrast to the edges of the tiles. This confounding factor makes it difficult to relate the observed effect to a specific characteristic of the visual stimuli.

4.3. Modulation of frequency-specific power

Time-frequency analysis also provides great insight about the dynamic nature of V1 activity in response to visual stimulation. Although the t-test between scrambled and intact conditions does not result in many significant spots within the time-frequency map, several important findings do emerge. The most dominant findings are observed in the 'long' (400ms) presentation conditions, whereby a higher gamma power (60-80Hz) is observed in the scrambled versus the intact image condition. Further, this higher gamma power can be explained with a higher prediction error. Recent findings have shown that bottom-up prediction errors propagate forward on a gamma frequency channel (Arnal et al., 2011; Bastos et al., 2015; Bauer et al., 2014; van Kerkoerle et al., 2014). Similarly, our findings are consistent with the results reported by an MEG study that used audio-visual stimuli (Arnal et al., 2011). Under the use of these multi-sensory stimuli, the visual input may or may not correctly predict the following auditory input. When the visual input did not foretell the auditory stimulus, higher gamma power in lower sensory areas was reported (Arnal et al., 2011). Similarly, this effect is observed in this study under 'long' presentation conditions, specifically from 100-400ms post-stimulation. Thus, these results parallel the significant difference observed in the VEP of area V1 (Fig. 3-13(a)).

In general, there are few studies that have used natural images and their phase-scrambled counterparts to investigate the modulation of activity in primary visual cortex. However, several

studies exist whereby similar stimuli were used to investigate the sensitivity of higher-order areas to objects, faces, or shapes (Henson et al., 2007; Henson et al., 2009; Meeren et al., 2013). These studies all report stronger VEP in response to intact images, as recorded by lateral occipital sensors. This activity is further localized in higher visual areas, such as LO and V4. Meeren et al. (2013) also computed the Global Field Power of the MEG sensor, similar to our sensor-level RMS analysis. They observed stronger magnetic field power at approximately 120ms post-stimulation in the phase-scrambled condition compared to the natural image condition. They also observe stronger response power to intact images after approximately 250ms post-stimulation (Meeren et al., 2013). Except for the precise timing of this effect, these results coincide with those of this thesis at the sensor-level.

Other than the Rieger et al. MEG and fMRI study, which was discussed in detail in this section, Hemond et al. (2007) and Malach et al. (1995) are two fMRI studies that use similar stimuli to investigate V1 activation. They also report an increase in the BOLD signal of area V1 in response to phase-scrambled images compared to natural images. (Hemond et al., 2007; Malach et al., 1995). If the linear relationship between VEP and BOLD signal is considered, according to (Obrig et al., 2002), the results of these studies are in accordance with the findings of this MEG experiment. In addition, several similar fMRI studies have employed different approaches to perturb the spatial coherence or structure of a visual input and observe the effect of this perturbation on V1 activity. Some of these approaches, such as masking images (Bar et al., 2006; Chen et al., 2010), tile-scrambling (Huberle and Lutzenberger, 2013), texture and contour extractions (Dumoulin et al., 2008) or the addition of graded levels of noise (Tjan et al., 2006), are less controllable than phase shuffling with respect to image statistics. Others used well

controlled stimuli but very different from what we employed in our study (Dumoulin et al., 2008; Freeman et al., 2013; Murray et al., 2002). It is interesting to note that some of these studies report an increased V1 BOLD signal as the visual input becomes less coherent or globally structured (Chen et al., 2010; Dumoulin and Hess, 2006; Murray et al., 2002). Yet, other studies fail to identify any modulation in V1 activity in response to changed level of naturalness or noise (Freeman et al., 2013; Tjan et al., 2006).

In summary, this thesis relates modulation of neuronal activity to image phase-scrambling, which is observed as early as 50ms post-stimulation. There are several fMRI and MEG/EEG studies that either fail to detect this modulatory effect (Dumoulin et al., 2008; Freeman et al., 2013; Rieger et al., 2013; Tjan et al., 2006) or report it only in later stages of processing (after 100ms) (Rieger et al., 2013). However, special factors accounted for by my study, as summarized below, account for this difference. Firstly, the stimulus was designed in a careful manner. The RMS and mean luminance normalization, the image edge smoothing, and the placement of images in one quadrant of the visual field reduce the confounding factors in the observed results. Moreover, in contrast to other methods of image scrambling (for instance, tile-scrambling), the method employed in this study is highly controllable in terms of different statistics of the image. Further, the careful attention binding task and its subsequent data analysis diminish the effect of attention on the obtained results. In contrast to the majority of MEG/EEG studies, the data analysis was not limited to the sensor space. The neuronal activation was projected onto the cortical surface via source reconstruction, enabling the investigation of precise cortical areas like primary visual cortex (V1). Finally, unlike previous studies that addressed similar questions, individual

differences were accounted for during the processes of ROI definition and temporal response alignment.

5. Conclusion and future work

This study reports modulated V1 activity associated with changes in the higher-order statistics of the visual input. These findings are consistent with models of visual information processing that recognize feedback modulation in lower visual areas.

The results are further improved by more precisely localizing this activation in the source space, which can be achieved via the use of fMRI priors for source localization and requires applying the same paradigm in an MRI scanner. Moreover, the interaction between different visual areas can be assessed using DCM or other functional connectivity metrics. In doing so, further insight may be obtained regarding the flow of information between visual areas in response to natural and artificial texture image stimulation. Nonlinear mapping of individual VEPs to a standard template may be a more practical solution than the simple alignment approach used in this experiment. However, this approach would require more careful interpretation.

6. References

[1] http://www.cambridgeincolour.com/tutorials/gamma-correction.htm.

[2] http://www.freeimages.com/.

[3] www.techcyn.com/feature.php?id=f2&issue=1.

[4] Ales, J., Carney, T., & Klein, S. A. (2010). The folding fingerprint of visual cortex reveals the timing of human V1 and V2. *Neuroimage, 49*(3), 2494-2502. doi: 10.1016/j.neuroimage.2009.09.022

[5] Ales, J. M., Yates, J. L., & Norcia, A. M. (2013). On determining the intracranial sources of visual evoked potentials from scalp topography: a reply to Kelly et al. (this issue). *Neuroimage, 64*, 703-711. doi: 10.1016/j.neuroimage.2012.09.009

[6] Alink, A., Schwiedrzik, C. M., Kohler, A., Singer, W., & Muckli, L. (2010). Stimulus predictability reduces responses in primary visual cortex. *J Neurosci, 30*(8), 2960-2966. doi: 10.1523/JNEUROSCI.3730-10.2010

[7] Arnal, L. H., Wyart, V., & Giraud, A. L. (2011). Transitions in neural oscillations reflect prediction errors generated in audiovisual speech. *Nat Neurosci, 14*(6), 797-801. doi: 10.1038/nn.2810

[8] Attias, H., and Christoph E. Schreiner. (1998). Coding of naturalistic stimuli by auditory midbrain neurons. *Advances in neural information processing systems*, 103-109.

[9] Attneave, F. (1954). Some informational aspects of visual perception. *Psychol Rev, 61*(3), 183-193.

[10] Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *J Cogn Neurosci, 15*(4), 600-609. doi: 10.1162/089892903321662976

[11] Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., . . . Halgren, E. (2006). Top-down facilitation of visual recognition. *Proc Natl Acad Sci U S A, 103*(2), 449-454. doi: 10.1073/pnas.0507062103

[12] Barlow, H. B. (1961). Possible principles underlying the transformation of sensory messages. *Sensory communication*, 217-234.

[13] Bartels, A., & Zeki, S. (2005). The chronoarchitecture of the cerebral cortex. *Philos Trans R Soc Lond B Biol Sci, 360*(1456), 733-750. doi: 10.1098/rstb.2005.1627

[14] Bastos, A. M., Vezoli, J., Bosman, C. A., Schoffelen, J. M., Oostenveld, R., Dowdall, J. R., . . . Fries, P. (2015). Visual Areas Exert Feedforward and Feedback Influences through Distinct Frequency Channels. *Neuron, 85*(2), 390-401. doi: 10.1016/j.neuron.2014.12.018

[15] Bauer, M., Stenner, M. P., Friston, K. J., & Dolan, R. J. (2014). Attentional modulation of alpha/beta and gamma oscillations reflect functionally distinct processes. *J Neurosci, 34*(48), 16117-16125. doi: 10.1523/JNEUROSCI.3474-13.2014

[16] Betti, V., Della Penna, S., de Pasquale, F., Mantini, D., Marzetti, L., Romani, G. L., & Corbetta, M. (2013). Natural scenes viewing alters the dynamics of functional connectivity in the human brain. *Neuron, 79*(4), 782-797. doi: 10.1016/j.neuron.2013.06.022

[17] Camprodon, J. A., Zohary, E., Brodbeck, V., & Pascual-Leone, A. (2010). Two phases of V1 activity for visual recognition of natural images. *J Cogn Neurosci, 22*(6), 1262-1269. doi: 10.1162/jocn.2009.21253

[18] Chao, L. L., Haxby, J. V., & Martin, A. (1999). Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nat Neurosci, 2*(10), 913-919. doi: 10.1038/13217

[19] Chaumon, M., Drouet, V., & Tallon-Baudry, C. (2008). Unconscious associative memory affects visual

processing before 100 ms. *J Vis, 8*(3), 10 11-10. doi: 10.1167/8.3.10

[20] Chen, J., Zhou, T., Yang, H., & Fang, F. (2010). Cortical dynamics underlying face completion in human visual system. *J Neurosci, 30*(49), 16692-16698. doi: 10.1523/JNEUROSCI.3578-10.2010

[21] Codispoti, M., Ferrari, V., Junghofer, M., & Schupp, H. T. (2006). The categorization of natural scenes: brain attention networks revealed by dense sensor ERPs. *Neuroimage, 32*(2), 583-591. doi: 10.1016/j.neuroimage.2006.04.180

[22] den Ouden, H. E., Daunizeau, J., Roiser, J., Friston, K. J., & Stephan, K. E. (2010). Striatal prediction error modulates cortical coupling. *J Neurosci, 30*(9), 3210-3219. doi: 10.1523/JNEUROSCI.4458-09.2010

[23] den Ouden, H. E., Friston, K. J., Daw, N. D., McIntosh, A. R., & Stephan, K. E. (2009). A dual role for prediction error in associative learning. *Cereb Cortex, 19*(5), 1175-1185. doi: 10.1093/cercor/bhn161

[24] Di Russo, F., Pitzalis, S., Spitoni, G., Aprile, T., Patria, F., Spinelli, D., & Hillyard, S. A. (2005). Identification of the neural sources of the pattern-reversal VEP. *Neuroimage, 24*(3), 874-886. doi: 10.1016/j.neuroimage.2004.09.029

[25] Di Russo, F., Stella, A., Spitoni, G., Strappini, F., Sdoia, S., Galati, G., . . . Pitzalis, S. (2012). Spatiotemporal brain mapping of spatial attention effects on pattern-reversal ERPs. *Hum Brain Mapp, 33*(6), 1334-1351. doi: 10.1002/hbm.21285

[26] Dumoulin, S. O., Dakin, S. C., & Hess, R. F. (2008). Sparsely distributed contours dominate extra-striate responses to complex scenes. *Neuroimage, 42*(2), 890-901. doi: 10.1016/j.neuroimage.2008.04.266

[27] Dumoulin, S. O., & Hess, R. F. (2006). Modulation of V1 activity by shape: image-statistics or shape-based perception? *J Neurophysiol, 95*(6), 3654-3664. doi: 10.1152/jn.01156.2005

[28] Egner, T., Monti, J. M., & Summerfield, C. (2010). Expectation and surprise determine neural population responses in the ventral visual stream. *J Neurosci, 30*(49), 16601-16608. doi: 10.1523/JNEUROSCI.2770-10.2010

[29] Felsen, G., & Dan, Y. (2005). A natural approach to studying vision. *Nat Neurosci, 8*(12), 1643-1646. doi: 10.1038/nn1608

[30] Felsen, G., Touryan, J., Han, F., & Dan, Y. (2005). Cortical sensitivity to visual features in natural scenes. *PLoS Biol, 3*(10), e342. doi: 10.1371/journal.pbio.0030342

[31] Forman, S. D., Cohen, J. D., Fitzgerald, M., Eddy, W. F., Mintun, M. A., & Noll, D. C. (1995). Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. *Magn Reson Med, 33*(5), 636-647.

[32] Foxe, J. J., & Simpson, G. V. (2002). Flow of activation from V1 to frontal cortex in humans. A framework for defining "early" visual processing. *Exp Brain Res, 142*(1), 139-150. doi: 10.1007/s00221-001-0906-7

[33] Freeman, J., & Simoncelli, E. P. (2011). Metamers of the ventral stream. *Nat Neurosci, 14*(9), 1195-1201. doi: 10.1038/nn.2889

[34] Freeman, J., Ziemba, C. M., Heeger, D. J., Simoncelli, E. P., & Movshon, J. A. (2013). A functional and perceptual signature of the second visual area in primates. *Nat Neurosci, 16*(7), 974-981. doi: 10.1038/nn.3402

[35] Froudarakis, E., Berens, P., Ecker, A. S., Cotton, R. J., Sinz, F. H., Yatsenko, D., . . . Tolias, A. S. (2014). Population code in mouse V1 facilitates readout of natural scenes through increased sparseness. *Nat Neurosci*. doi: 10.1038/nn.3707

[36] Fukushima, K. (1980). Neocognitron: a self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol Cybern, 36*(4), 193-202.

[37] Garrido, M. I., Kilner, J. M., Kiebel, S. J., & Friston, K. J. (2007). Evoked brain responses are generated by

feedback loops. *Proc Natl Acad Sci U S A, 104*(52), 20961-20966. doi: 10.1073/pnas.0706274105

[38] Graham, N. (1980). Spatial-frequency channels in human vision: detecting edges without edge detectors. *Visual coding and adaptability 3*, 215-262.

[39] Grill-Spector, K., Kushnir, T., Hendler, T., Edelman, S., Itzchak, Y., & Malach, R. (1998). A sequence of object-processing stages revealed by fMRI in the human occipital lobe. *Hum Brain Mapp, 6*(4), 316-328.

[40] Groen, Ii, Ghebreab, S., Prins, H., Lamme, V. A., & Scholte, H. S. (2013). From image statistics to scene gist: evoked neural activity reveals transition from low-level natural image structure to scene category. *J Neurosci, 33*(48), 18814-18824. doi: 10.1523/JNEUROSCI.3128-13.2013

[41] Hagler, D. J., Jr., Halgren, E., Martinez, A., Huang, M., Hillyard, S. A., & Dale, A. M. (2009). Source estimates for MEG/EEG visual evoked responses constrained by multiple, retinotopically-mapped stimulus locations. *Hum Brain Mapp, 30*(4), 1290-1309. doi: 10.1002/hbm.20597

[42] Hansen, B. C., & Hess, R. F. (2007). Structural sparseness and spatial phase alignment in natural scenes. *J Opt Soc Am A Opt Image Sci Vis, 24*(7), 1873-1885.

[43] Hemond, C. C., Kanwisher, N. G., & Op de Beeck, H. P. (2007). A preference for contralateral stimuli in human object- and face-selective cortex. *PLoS One, 2*(6), e574. doi: 10.1371/journal.pone.0000574

[44] Henriksson, L., Hyvarinen, A., & Vanni, S. (2009). Representation of cross-frequency spatial phase relationships in human visual cortex. *J Neurosci, 29*(45), 14342-14351. doi: 10.1523/JNEUROSCI.3136-09.2009

[45] Henson, R. N., Goshen-Gottstein, Y., Ganel, T., Otten, L. J., Quayle, A., & Rugg, M. D. (2003). Electrophysiological and haemodynamic correlates of face perception, recognition and priming. *Cereb Cortex, 13*(7), 793-805.

[46] Henson, R. N., Mattout, J., Singh, K. D., Barnes, G. R., Hillebrand, A., & Friston, K. (2007). Population-level inferences for distributed MEG source localization under multiple constraints: application to face-evoked fields. *Neuroimage, 38*(3), 422-438. doi: 10.1016/j.neuroimage.2007.07.026

[47] Henson, R. N., Mouchlianitis, E., & Friston, K. J. (2009). MEG and EEG data fusion: simultaneous localisation of face-evoked responses. *Neuroimage, 47*(2), 581-589. doi: 10.1016/j.neuroimage.2009.04.063

[48] Huberle, E., & Lutzenberger, W. (2013). Temporal properties of shape processing by event-related MEG adaptation. *Neuroimage, 67*, 119-126. doi: 10.1016/j.neuroimage.2012.10.070

[49] Hubner, M., Caelli, T., & Rentschler, I. (1988). Visual phase resolution for gray-scale textures. *Percept Psychophys, 43*(4), 319-325.

[50] Kelly, S. P., Gomez-Ramirez, M., & Foxe, J. J. (2008). Spatial attention modulates initial afferent activity in human primary visual cortex. *Cereb Cortex, 18*(11), 2629-2636. doi: 10.1093/cercor/bhn022

[51] Kelly, S. P., Vanegas, M. I., Schroeder, C. E., & Lalor, E. C. (2013). The cruciform model of striate generation of the early VEP, re-illustrated, not revoked: a reply to Ales et al. (2013). *Neuroimage, 82*, 154-159. doi: 10.1016/j.neuroimage.2013.05.112

[52] Knill, D. C., and David Field. (1990). Human discrimination of fractal images. *JOSA A 7, 6*, 1113-1123.

[53] Kok, P., Jehee, J. F., & de Lange, F. P. (2012). Less is more: expectation sharpens representations in the primary visual cortex. *Neuron, 75*(2), 265-270. doi: 10.1016/j.neuron.2012.04.034

[54] Kok, P., Rahnev, D., Jehee, J. F., Lau, H. C., & de Lange, F. P. (2012). Attention reverses the effect of prediction in silencing sensory signals. *Cereb Cortex, 22*(9), 2197-2206. doi: 10.1093/cercor/bhr310

[55] Konig, S. D., & Buffalo, E. A. (2014). A nonparametric method for detecting fixations and saccades using cluster analysis: removing the need for arbitrary thresholds. *J Neurosci Methods, 227*, 121-131. doi: 10.1016/j.jneumeth.2014.01.032

[56] Larsson, J., & Heeger, D. J. (2006). Two retinotopic visual areas in human lateral occipital cortex. *J Neurosci, 26*(51), 13128-13142. doi: 10.1523/JNEUROSCI.1657-06.2006

[57] Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *J Opt Soc Am A Opt Image Sci Vis, 20*(7), 1434-1448.

[58] Lennie, P. (1998). Single units and visual cortical organization. *Perception, 27*(8), 889-935.

[59] Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. A., . . . Tootell, R. B. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proc Natl Acad Sci U S A, 92*(18), 8135-8139.

[60] Mantini, D., Corbetta, M., Romani, G. L., Orban, G. A., & Vanduffel, W. (2012). Data-driven analysis of analogous brain networks in monkeys and humans during natural vision. *Neuroimage, 63*(3), 1107-1118. doi: 10.1016/j.neuroimage.2012.08.042

[61] Martinez-Conde, S., Macknik, S. L., Troncoso, X. G., & Hubel, D. H. (2009). Microsaccades: a neurophysiological analysis. *Trends Neurosci, 32*(9), 463-475. doi: 10.1016/j.tins.2009.05.006

[62] Meeren, H. K., de Gelder, B., Ahlfors, S. P., Hamalainen, M. S., & Hadjikhani, N. (2013). Different cortical dynamics in face and body perception: an MEG study. *PLoS One, 8*(9), e71408. doi: 10.1371/journal.pone.0071408

[63] Meeren, H. K., Hadjikhani, N., Ahlfors, S. P., Hamalainen, M. S., & de Gelder, B. (2008). Early category-specific cortical activation revealed by visual stimulus inversion. *PLoS One, 3*(10), e3503. doi: 10.1371/journal.pone.0003503

[64] Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: two cortical pathways. *Trends Neurosci, 6*, 414-417.

[65] Miyashita, Y., & Hayashi, T. (2000). Neural representation of visual objects: encoding and top-down activation. *Curr Opin Neurobiol, 10*(2), 187-194.

[66] Mumford, D. (1992). On the computational architecture of the neocortex. II. The role of cortico-cortical loops. *Biol Cybern, 66*(3), 241-251.

[67] Murray, S. O., Kersten, D., Olshausen, B. A., Schrater, P., & Woods, D. L. (2002). Shape perception reduces activity in human primary visual cortex. *Proc Natl Acad Sci U S A, 99*(23), 15164-15169. doi: 10.1073/pnas.192579399

[68] Naselaris, T., Prenger, R. J., Kay, K. N., Oliver, M., & Gallant, J. L. (2009). Bayesian reconstruction of natural images from human brain activity. *Neuron, 63*(6), 902-915. doi: 10.1016/j.neuron.2009.09.006

[69] Nishimoto, S., Vu, A. T., Naselaris, T., Benjamini, Y., Yu, B., & Gallant, J. L. (2011). Reconstructing visual experiences from brain activity evoked by natural movies. *Curr Biol, 21*(19), 1641-1646. doi: 10.1016/j.cub.2011.08.031

[70] Obrig, H., Israel, H., Kohl-Bareis, M., Uludag, K., Wenzel, R., Muller, B., . . . Villringer, A. (2002). Habituation of the visually evoked potential and its vascular response: implications for neurovascular coupling in the healthy adult. *Neuroimage, 17*(1), 1-18.

[71] Olshausen, B. A. (2013). Perception as an Inference Problem.

[72] Olshausen, B. A., & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature, 381*(6583), 607-609. doi: 10.1038/381607a0

[73] Olshausen, B. A., & Field, D. J. (1996). Natural image statistics and efficient coding. *Network, 7*(2), 333-339. doi: 10.1088/0954-898X/7/2/014

[74] Olshausen, B. A., & Field, D. J. (1997). Sparse coding with an overcomplete basis set: a strategy employed by V1? *Vision Res, 37*(23), 3311-3325.

[75] Ossandon, J. P., Helo, A. V., Montefusco-Siegmund, R., & Maldonado, P. E. (2010). Superposition model predicts EEG occipital activity during free viewing of natural scenes. *J Neurosci, 30*(13), 4787-4795. doi: 10.1523/JNEUROSCI.5769-09.2010

[76] Parraga, C. A., Tom Troscianko, David J. Tolhurst. (2000). The human visual system is optimised for processing the spatial information in natural visual images. *Current Biology 10, 1*, 35-38.

[77] Poghosyan, V., & Ioannides, A. A. (2008). Attention modulates earliest responses in the primary auditory and visual cortices. *Neuron, 58*(5), 802-813. doi: 10.1016/j.neuron.2008.04.013

[78] Rainer, G., Augath, M., Trinath, T., & Logothetis, N. K. (2001). Nonmonotonic noise tuning of BOLD fMRI signal to natural images in the visual cortex of the anesthetized monkey. *Curr Biol, 11*(11), 846-854.

[79] Rainer, G., Augath, M., Trinath, T., & Logothetis, N. K. (2002). The effect of image scrambling on visual cortical BOLD activity in the anesthetized monkey. *Neuroimage, 16*(3 Pt 1), 607-616.

[80] Ranganath, C., Cohen, M. X., Dam, C., & D'Esposito, M. (2004). Inferior temporal, prefrontal, and hippocampal contributions to visual working memory maintenance and associative memory retrieval. *J Neurosci, 24*(16), 3917-3925. doi: 10.1523/JNEUROSCI.5053-03.2004

[81] Rao, H., Wang, J., Tang, K., Pan, W., & Detre, J. A. (2007). Imaging brain activity during natural vision using CASL perfusion fMRI. *Hum Brain Mapp, 28*(7), 593-601. doi: 10.1002/hbm.20288

[82] Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat Neurosci, 2*(1), 79-87. doi: 10.1038/4580

[83] Rauss, K., Schwartz, S., & Pourtois, G. (2011). Top-down effects on early visual processing in humans: a predictive coding framework. *Neurosci Biobehav Rev, 35*(5), 1237-1253. doi: 10.1016/j.neubiorev.2010.12.011

[84] Rieger, J. W., Gegenfurtner, K. R., Schalk, F., Koechy, N., Heinze, H. J., & Grueschow, M. (2013). BOLD responses in human V1 to local structure in natural scenes: Implications for theories of visual coding. *J Vis, 13*(2), 19. doi: 10.1167/13.2.19

[85] Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nat Neurosci, 2*(11), 1019-1025. doi: 10.1038/14819

[86] S.C., D. What causes non-monotonic tuning of fMRI response to noisy images?

[87] Sayres, R., & Grill-Spector, K. (2008). Relating retinotopic and object-selective responses in human lateral occipital cortex. *J Neurophysiol, 100*(1), 249-267. doi: 10.1152/jn.01383.2007

[88] Scholte, H. S., Ghebreab, S., Waldorp, L., Smeulders, A. W., & Lamme, V. A. (2009). Brain responses strongly correlate with Weibull image statistics when processing natural images. *J Vis, 9*(4), 29 21-15. doi: 10.1167/9.4.29

[89] Schwarzlose, R. F., Swisher, J. D., Dang, S., & Kanwisher, N. (2008). The distribution of category and location information across object-selective regions in human visual cortex. *Proc Natl Acad Sci U S A, 105*(11), 4447-4452. doi: 10.1073/pnas.0800431105

[90] Simoncelli, E. P., & Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annu Rev Neurosci, 24*, 1193-1216. doi: 10.1146/annurev.neuro.24.1.1193

[91] Sohoglu, E., Peelle, J. E., Carlyon, R. P., & Davis, M. H. (2012). Predictive top-down integration of prior knowledge during speech perception. *J Neurosci, 32*(25), 8443-8453. doi: 10.1523/JNEUROSCI.5069-11.2012

[92] Srinivasan, M. V., Laughlin, S. B., & Dubs, A. (1982). Predictive coding: a fresh view of inhibition in the retina. *Proc R Soc Lond B Biol Sci, 216*(1205), 427-459.

[93] Sun, J., Hu, X., Huang, X., Liu, Y., Li, K., Li, X., . . . Zhang, J. (2012). Inferring consistent functional interaction patterns from natural stimulus FMRI data. *Neuroimage, 61*(4), 987-999. doi: 10.1016/j.neuroimage.2012.01.142

[94] Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D., & Leahy, R. M. (2011). Brainstorm: a user-friendly application for MEG/EEG analysis. *Comput Intell Neurosci, 2011*, 879716. doi: 10.1155/2011/879716

[95] Tanaka, K. (1993). Neuronal mechanisms of object recognition. *Science, 262*(5134), 685-688.

[96] Tanskanen, T., Nasanen, R., Montez, T., Paallysaho, J., & Hari, R. (2005). Face recognition and cortical responses show similar sensitivity to noise spatial frequency. *Cereb Cortex, 15*(5), 526-534. doi: 10.1093/cercor/bhh152

[97] Thomson, M. G., Foster, D. H., & Summers, R. J. (2000). Human sensitivity to phase perturbations in natural images: a statistical framework. *Perception, 29*(9), 1057-1069.

[98] Tjan, B. S., Lestou, V., & Kourtzi, Z. (2006). Uncertainty and invariance in the human visual cortex. *J Neurophysiol, 96*(3), 1556-1568. doi: 10.1152/jn.01367.2005

[99] Todorovic, A., & de Lange, F. P. (2012). Repetition suppression and expectation suppression are dissociable in time in early auditory evoked fields. *J Neurosci, 32*(39), 13389-13395. doi: 10.1523/JNEUROSCI.2227-12.2012

[100] Tomita, H., Ohbayashi, M., Nakahara, K., Hasegawa, I., & Miyashita, Y. (1999). Top-down signal from prefrontal cortex in executive control of memory retrieval. *Nature, 401*(6754), 699-703. doi: 10.1038/44372

[101] van Kerkoerle, T., Self, M. W., Dagnino, B., Gariel-Mathis, M. A., Poort, J., van der Togt, C., & Roelfsema, P. R. (2014). Alpha and gamma oscillations characterize feedback and feedforward processing in monkey visual cortex. *Proc Natl Acad Sci U S A, 111*(40), 14332-14341. doi: 10.1073/pnas.1402773111

[102] Vinje, W. E., & Gallant, J. L. (2000). Sparse coding and decorrelation in primary visual cortex during natural vision. *Science, 287*(5456), 1273-1276.

[103] Wacongne, C., Changeux, J. P., & Dehaene, S. (2012). A neuronal model of predictive coding accounting for the mismatch negativity. *J Neurosci, 32*(11), 3665-3678. doi: 10.1523/JNEUROSCI.5003-11.2012

[104] Wacongne, C., Labyt, E., van Wassenhove, V., Bekinschtein, T., Naccache, L., & Dehaene, S. (2011). Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proc Natl Acad Sci U S A, 108*(51), 20754-20759. doi: 10.1073/pnas.1117807108

[105] Wallis, G., & Rolls, E. T. (1997). Invariant face and object recognition in the visual system. *Prog Neurobiol, 51*(2), 167-194.

[106] Wang, X. J. (2010). Neurophysiological and computational principles of cortical rhythms in cognition. *Physiol Rev, 90*(3), 1195-1268. doi: 10.1152/physrev.00035.2008

[107] Whittingstall, K., Bartels, A., Singh, V., Kwon, S., & Logothetis, N. K. (2010). Integration of EEG source imaging and fMRI during continuous viewing of natural movies. *Magn Reson Imaging, 28*(8), 1135-1142. doi: 10.1016/j.mri.2010.03.042

[108] Woodhead, Z. V., Barnes, G. R., Penny, W., Moran, R., Teki, S., Price, C. J., & Leff, A. P. (2014). Reading front to back: MEG evidence for early feedback effects during word recognition. *Cereb Cortex, 24*(3), 817-825. doi: 10.1093/cercor/bhs365

7. Appendix

7.1. fMRI scan: stimulus parameters and data acquisition

In the fMRI scan for the retinotopic mapping, the TR was set to 2 seconds and it was equal to the duration of presentation of each ring/wedge at each location. Each cycle of the stimulus was 30 seconds long, and each run consisted of 10 cycles. In one fMRI session I embedded 8 runs which were divided equally into 4 paradigms: expanding and contracting rings, and clock-wise and counter-clock-wise rotating wedges.

In Fig. 2-12 one can see the stimulus I used to delineate the motion sensitive area MT. It consists of 32s cycles of full-field static and moving dots patterns on a gray background. The dots appeared randomly in the visual field and moved towards or away from the fixation point. I ran this stimulus 8 cycles per run and 2 runs per subjects in each fMRI session.

I also designed and implemented an LO localizer using a database of 270 gray-scale images of objects and scenes. I computed the phase randomized version of these images using the same method explained in Section 2.1.2. In Fig. 2-12(b) one can see the screenshots of the LO localizer stimulus which consists of 24s periods of intact images (12s) following by their phase-scrambled versions (12s) matching in contrast and mean luminance. Each single image lasts on the screen for 250ms (the parameters are the same as in (Larsson and Heeger, 2006)).

For the retinotopy and MT-localizer, I modified stimuli that were used previously in our Lab. I scanned two subjects using this paradigm. The distance from the eyes to the mirror was 12.5cm and from the mirror to the screen was 134.5cm. The field of view was 24°×18° in width and height respectively. In all runs, subjects were asked to respond to the color change of a small fixation point by pressing a button. The changes took place randomly at intervals in the range of 3 to 6 seconds. Two runs of T1 MR Scan were also obtained for each subject within the same session. The data were acquired using a 3 Tesla MR scanner at Montreal Neurological Institute Brain imaging Centre.

I performed the retinotopy analysis using Freesurfer v5.1. Firstly, I reconstructed the brain surface from the data averaged over the two structural runs. Next, I specified the properties of each run (the stimulus type and the direction of motion) in .par files. I registered the anatomy and BOLD scans using the bbregister command in Freesurfer and applied the default pre-processing stages on each run. Finally, I conducted the retinotopy analysis in Freesurfer and projected the maps on the flattened surface of the occipital lope. For the MT-localizer data, I applied a contrast analysis in Freesurfer between static and moving dots conditions.