Redundancy in the visual system

Redundancy in the visual system

Kirsten Petras

Thèse présentée en vue de l'obtention du grade de Docteur en sciences psychologiques et de l'éducation

Promoter

Prof. Dr. Valerie Goffaux (UC Louvain, Belgium)

Accompanying committee

Prof. Dr. Bruno Rossion (Université de Lorraine, France)
Prof. Dr. André Mouraux, MD (UC Louvain, Belgium)

UCL PRESSES UNIVERSITAIRES DE LOUVAIN

Collection de thèses de l'Université catholique de Louvain, 2020

President

Prof. Dr. André Mouraux, MD (UC Louvain, Belgium)

Accompanying committee and jury

Prof. Dr. Valerie Goffaux (UC Louvain, Belgium)
Prof. Dr. Tallon-Baudry (Ecole Normale Supérieure, France)
Prof. Dr. Sarang S. Dalal (Aarhus University, Denmark)
Prof. Dr. Gilles Vannuscorps (UC Louvain, Belgium)
Prof. Dr. Bruno Rossion (Université de Lorraine, France)
Prof. Dr. André Mouraux, MD (UC Louvain, Belgium)

Cover: Dennis Croonenberg & Kirsten Petras

Acknowledgments

Hopefully, after a successful defense, this thesis will lead me to be awarded the academic degree of doctor. How strange. Many people contributed to every aspect of the work presented in these pages, yet only one person gets a piece of paper certifying this communal effort. I want to take the opportunity to acknowledge my colleagues, friends, and family without whom, I am certain, this thesis would not have happened.

First and foremost, I want to thank my supervisor, Valerie Goffaux for taking a chance on me when I was almost ready to give up and then putting up with me over five years of sometimes heated discussions and constant begging for a pony (which, I would like to point out, I still didn't get). Thank you for your unwavering support and for creating a lab environment where integrity and collaboration are valued.

Thanks to my accompanying committee, Bruno Rossion and André Mouraux for their support. I appreciate the time you took out of your busy schedules to discuss with me or put me into contact with people to discuss with.

Sarang, thank you for welcoming me into your lab. I had a wonderful time in Aarhus and learned a lot from you and your lab members.

Alex and Jordan, thank you for giving me refuge in your apartments and for many wine and gin tonic fueled discussions.

My lab colleagues, Jolien, Matthew, Yu Fang and Umut: Thanks for the many lunches over strange topics of conversation and the dead pigeon updates when I wasn't in LLN to witness the decay of the latest one live in front of the office window. And special thanks, Jolien and Yu Fang, for the many times you restarted my computer after I broke it while working remotely from Maastricht.

Christianne, all the podcasts in the world cannot replace you as a driving buddy. You have gone from being the world's best duckie to a dear friend.

All the "Italians", Mohammed, Jyothi, Francesca, Federica, Remi, Ceren, Roberta, Siddharth, Marco, Ane and Stefania: Thank you for making the sun shine in Louvain la Neuve.

Joan, I hope life outside academia treats you well and thank you, as well as Corentin, Milena and Aliette, for your patience with all the loud gatherings happening in the kitchen and in front of your offices. Alice, it was a pleasure sharing the office with you and I hope post-PhD life gives you everything you wish for.

The Maastricht people, Miriam, Yawen, Tahnee, Felix, Faruk, Fren, Rob, Chris, Esther and all the other people that are, or have been, in and around the CN department: You have made me feel welcome and I enjoyed every coffee break I shared with you (at this point, I also want to thank the great Bandito staff that provided the best coffee and cookies to go with it). Fabian, had me invade your office on a regular basis and still were never short on silly animal videos. Thank you for that! You also deserve a medal for the amount of experiments you participate in. Sri, I admire your calm kindness and your skill when it comes to MRI recording, analysis, and explaining all of that to others. I hope you and Chaitra have a wonderful time in Canada.
Sanne, where do I even start? From supervising my Master thesis project, introducing me to Matlab, making the contact with Valerie, helping me with all sorts of programming issues, collaborating in research and spending many afternoons working together at Zuid, I think it's fair to say that along with Valerie, you had the biggest influence on my scientific development. Thank you for that! But much more importantly, over this time you became a trusted friend and I thoroughly enjoy all the dinners, wijntjes and gardening projects we have together.
Mehrdad, Shyam and Lindsey, Nitzan and Wilbi, you are what made moving to Maastricht some years ago a good decision. Thank you! I hope we will enjoy many more random dinners with even more random topics of conversation together.

Dennis, ik find je heel lief! Thank you for being a true partner in the best possible sense. You never complain about me working too much (unless it's starting random building projects that then remain unfinished). You support me when I need support and call me out on my nonsense when it needs calling out. But most importantly, you make me feel loved every day. Thank you!

Mama, Papa, Miriam und Dominik, danke, dass ihr mich all die Jahre unterstützt habt und ihr heute noch jederzeit für mich da seid. Ihr seid die Besten!

CONTENTS

Chapter 1

General Introduction

Visual perception and statistical regularities in the visual diet

Imagine walking outside on a snowy day. Even with thick snowflakes falling past between you and a friend walking with you, you can still recognize her face. You can still 'see' the street, the houses to the side and the cars parked along the curb. This is possible because rather than being a faithful representation of the physical world around us, our visual system efficiently samples some of the available visual information and integrates it with prior knowledge about the statistical regularities of its environment, predictions, and flat out guesses to form what we experience as our visual perception.

A purely feedforward architecture, were a light-triggered response from the retina is converted to the final visual experience along successive brain areas, would not readily permit for the incorporation of information that is not directly contained within the visual stimulus. Thus, there must be some form of recurrence or top-down influence on visual processing. Although past research has made great progress in uncovering the dynamics of feedforward and feedback signaling along the visual pathways, how exactly the brain integrates the information available to it remains largely unknown. The work in this thesis aims to contribute to a mechanistic account of such information integration.

In the following I will argue that the statistical regularities of natural images are suitable to guide visual processing via feedback. For this I will first briefly introduce the concepts of feedforward and feedback processing in the visual system and then discuss how image redundancies relate to efficient neural representations.

Feedback and the ventral visual pathway

The basic anatomy of the visual system has been established through staining and tracing work and is remarkably similar across a wide range of species. In the feedforward direction, the signal generated by the photoreceptors in the retina travels through the bipolar and amacrine cells via the retinal ganglion to the optic

nerve, passes the optic chiasm where the nasal fibers from both eyes cross over and continues along the optic tract to the lateral geniculate nucleus (LGN). From there, the visual input reaches its first cortical destination which is the primary visual cortex (V1). Past V1, two main pathways emerge (Milner & Goodale, 1995; Mishkin, Ungerleider, & Macko, 1983). The dorsal, 'vision for action' stream has been linked to visually guided reach, the processing of visual motion and spatial location. It consists of a lateral and a medial group of regions, each of which receives separate projections from V1. The ventral visual pathway, also referred to as the "vision for perception" stream, continues on from V1 to V2, V4 and along the inferior temporal lobe. Along this pathway, receptive fields are becoming increasingly larger and the visual information they represent is becoming increasingly rich and invariant to more basic stimulus properties such as spatial frequency, line orientation, size and viewing angle (Hubel & Wiesel, 1977; Lennie, 1998; Maunsell & Newsome, 1987; Smith, Singh, Williams, & Greenlee, 2001; Zeki, 1978), until they are selective to complex categories such as faces, objects and scenes (Logothetis & Sheinberg, 1996; Ullman, 1996).

Whereas much of the increasing complexity and low-level invariance of receptive fields higher up this visual hierarchy can be explained by the feedforward convergence of preceding processing stages, many basic operations the visual system performs, such as figure ground segregation, i.e. the perceptual grouping of shapes as belonging to a visual object or the background, require information to flow from regions considered higher in the hierarchy to regions considered lower (Lee, Mumford, Romero, & Lamme, 1998; Roelfsema, Lamme, Spekreijse, & Bosch, 2002). Figure 1 illustrates that local edges and contrast boundaries are not sufficient to distinguish the figure from its background. However, V1 cells, that are initially sensitive to such local features, show a strongly enhanced response when their receptive field is on the figure rather than on the ground of an image. This enhancement can be observed from about 30-40ms after the cells initial response (Lamme, 1995) and cannot be explained by the cells intrinsic tuning properties, as the effect persists even if the figure defining boundary is well outside the cells receptive field. To explain modulations of V1 responses by stimulus information that is not inherently available to V1 receptive fields, lateral, recurrent and feedback processes need to be considered.

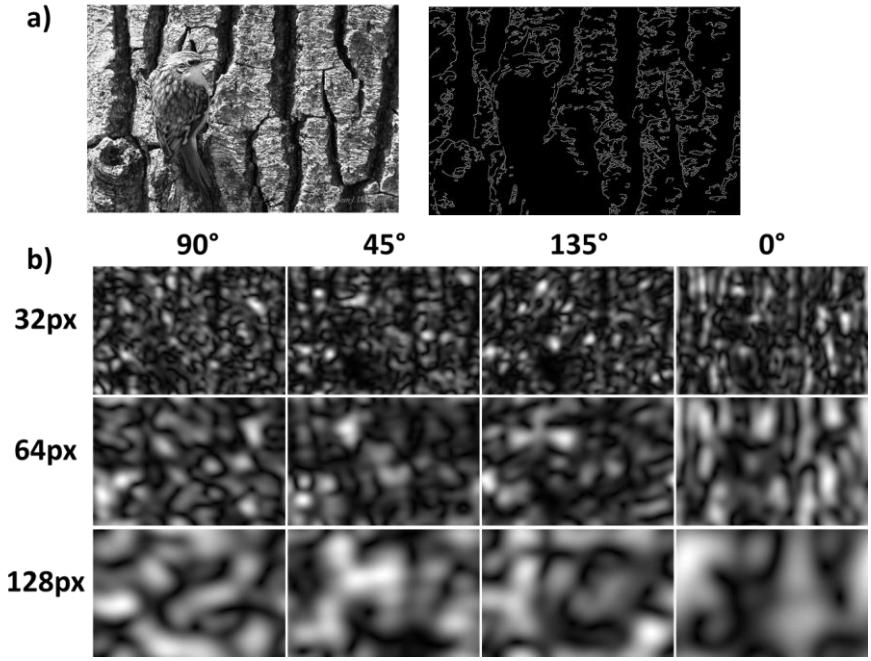


Figure 1: feedforward pooling of local filter properties is not sufficient to unambiguously segment the bird from the background. a) the canny edge algorithm fails to find the relevant boundaries of the figure. b) a bank of Gaussian filters with different orientation and spatial frequency properties is insufficient to extract the figure without further guidance.

Lateral, or horizontal connections are intra-areal influences that are thought to interact with feedback influences to enable e.g. contour integration (Liang et al., 2017). Recurrence refers to repetitive interareal feedforward-feedback loops (e.g. Neumann & Sepp, 1999), but is also sometimes used to describe intra-areal connections (Self, Kooijmans, Supèr, Lamme, & Roelfsema, 2012). The focus in the following is on interareal feedback processes, i.e. signaling from hierarchically higher to hierarchically lower visual areas.

The anatomical presence of cortical afferents to V1 has been uncontroversial even by the time Marr introduced his influential model of visual processing (Marr, 1982) almost 40 years ago and there is evidence for them from tracing studies (e.g. Burkhalter & Bernardo, 1989; Coogan & Burkhalter, 1990; Rockland & Virga, 1989; Van Essen, Anderson, & Felleman, 1992), animal electrophysiology(e.g. J Bullier, McCourt, & Henry, 1988; Gray, Engel, König, & Singer, 1990), human electrophysiology (Bastos et al., 2015; Kok, Failing, & de Lange, 2014; Michalareas

et al., 2016) and neuroimaging studies (e.g. Kok, Jehee, & De Lange, 2012; Muckli et al., 2015). What has been, and still is debated, is their operational relevance and their relation to V1 function.

The flow of information inconsistent with feedforward (here used to mean ascending the visual hierarchy as described above) in its various forms has been termed 'feedback' (J Bullier et al., 1988; Muckli et al., 2015), 'top down modulation' (Bar et al., 2006), 'bi-directional cortico-cortical connectivity' (Burkhalter & Bernardo, 1989), 'reverse hierarchy' (Ahissar & Hochstein, 1997, 2004), 'afferent connectivity' (Kennedy & Bullier, 1985), 'recurrent processing' (Neumann & Sepp, 1999; Self et al., 2012) or 're-entry' (Bouvier & Treisman, 2010) with the different terms sometimes used interchangeably. To complicate matters further, the term 're-entry' was initially coined by Gerald Edelman (1978) and was meant to describe parallel and recursive reciprocal signaling within and between brain areas. Re-entry according to Edelman is in stark contrast to feedback signaling as it shows no discernable directionality and no predefined input-output function (Edelman, 1993). Similar etymological confusion can be found for the expression 'top-down modulation' that is more commonly used to describe attentional modulation of perceptual processes that are not necessarily caused by prior visual stimulation (see for example Beck & Kastner, 2009; but also note Firestone & Scholl, 2016 for an alternative view). To disambiguate, here, 'feedback' is used to describe signaling from areas higher, to areas lower in the visual hierarchy, if the signal originally came from the area feedback is directed to[1]. Descending signals that do not directly rely on ascending visual input, such as attentional modulation or semantic context, are referred to as 'top-down' modulation.

Not only the terms describing descending information, but also the importance this information takes in models of visual perception differs across the literature. Marr acknowledged the existence of feedback connections throughout the visual system, but they do not play a critical role in his model. Instead, he proposed functional modules that perform distinct and progressively abstract operations through the feedforward convergence of local filter responses. In his view, V1 acts as a local feature detector and produces the "primal sketch" whereas extrastriate cortices

[1] Note that the very concept of a visual hierarchy is challenged when information transfer is seen as bi-directional (see e.g. Markov et al., 2014 or Hochstein & Ahissar, 2002). Nevertheless, the term is used here to refer to the order in which regions are reached via connections that target mid-laminar layers (e.g. Petro & Muckli, 2017)

such as V4 or IT respectively compute the 2½D sketch and 3D model (Marr, 1982). While Marr focused his work on a computational explanation of the steps necessary to form a coherent representation of the 3D world from a 2D projection onto the retina, Van Essen and Maunsell (1983) delivered a neurobiological model of a hierarchical visual system based on anatomical work. They note that cortico-cortical connections between visual areas are almost always reciprocal, thus if area A projects to area B, area A also receives projections from area B. They further note that the projections in one direction predominantly terminate in the granular layer (layer **IV**) of the receiving area, similar to the ascending thalamocortical pathways. By virtue of this similarity they term connections terminating in granular layers feedforward and connections terminating in supra- or infragranular layers feedback or descending connections. By laying out which areas project to which layer of which other areas, they arrive at a cortical hierarchy with areas receiving layer **IV** input being ranked above the area the input originated from (see figure 2).

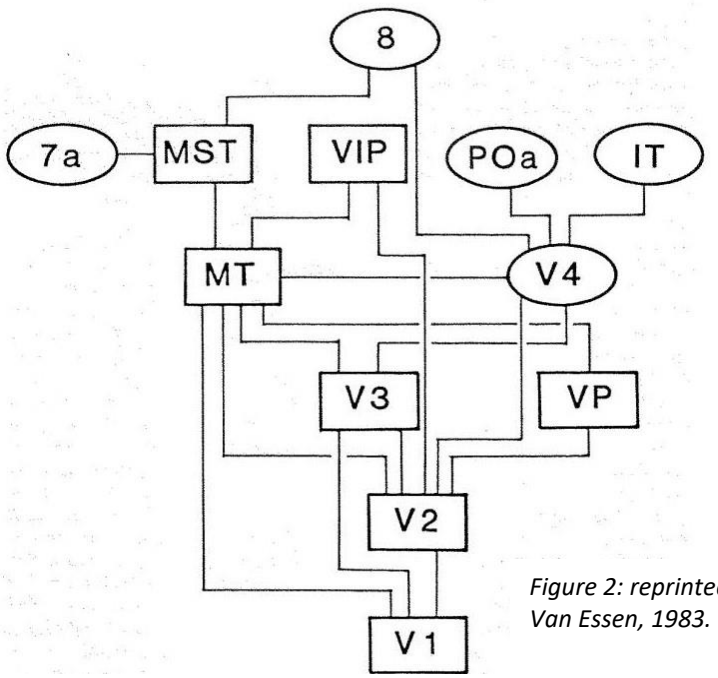


Figure 2: reprinted from Maunsell, Van Essen, 1983.

Drawing from Marr's computational approach to biological processing, Lee et al. (1998) formulated the idea that V1 neurons are changing their tuning properties over the course of visual processing and are intricately involved in higher level perceptual processing. In a series of experiments using single unit recordings in the V1 of awake, behaving monkeys they demonstrated that while V1 neurons initially responded to local features such as line orientation, their responses in a later time window were heavily influenced by contextual information (see figure 3).

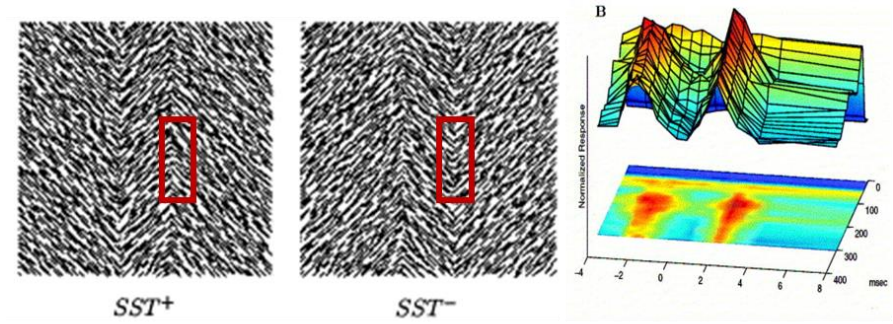


Figure 3: adapted from Lee et al., 1998. The left panel shows the stimuli used. The figure consists of a texture strip with the texture slanted 45° (SST^+) or 135° (SST^-) from the orientation of the boundary. Right panel: neurons with receptive fields parallel to the boundary (see red rectangle in left panel) showed an initial small response to the texture features in their non-preferred orientation. If the boundary defined by the slanted textures was in the cells preferred orientation, the boundary response that emerged after ~100ms was stronger than the initial 'local feature detector' response.

Importantly, this later stage response still reflected the neurons orientation preference, thus retaining the tuning properties of the early response period. When the boundary between a figure and the background was parallel to the neurons preferred orientation, the neuron, as expected, showed stronger responses when the boundary passed through its receptive field compared to when the figure or the background were placed on its receptive field. When textures defining the boundary did not have the cells preferred orientation, the response to the boundary in the later stage of processing far exceeded the initial response indicating a driving influence of figure properties outside the cells original receptive field. Those findings are inconsistent with a constant filter response to local features, as suggested by Marr, but instead support the conjecture that V1 neurons

serve as a “high resolution buffer *[…]* to perform all computations that integrate global information with spatial precision” (Lee et al., 1998). Jean Bullier (2001) extended those ideas in his integrated model of visual processing and refers to V1 and V2 as “active blackboards” that integrate information fed back from higher level visual regions. In line with the findings of Lee et al. he notes that a feedforward visual system in which the tuning properties of each stage depend solely on the combination of tuning properties of preceding stages, necessarily fails to segment objects from a visual scene once viewing conditions are even slightly suboptimal, may it be due to visual clutter, light and shade, partial occlusion or simply noisy images. Instead, an initial global analysis of the scene is necessary to guide the processing of local detail. Based on their very brief latencies of activation, on average within 70ms for area MT, Bullier identifies areas within the parietal cortex ‘dorsal stream’ as the most likely to influence responses in the inferior temporal cortex ‘ventral stream’ via the retro-injection of information into V1 and V2. Such feedback information could guide the high precision processing in early visual cortex to solve boundary problems and scene segmentation.

More recent characterizations of visual object recognition also suggest the involvement of ventral visual areas along the inferiotemporal cortex in the recurrent processing of visual stimuli. Kietzmann et al. (2019) have used source reconstructed MEG data in combination with deep neural network virtual cooling experiments to investigate the dynamics of visual object processing. Participants were shown images of animals, faces and inanimate objects while MEG was recorded. Subsequently, the authors constructed time resolved representational dissimilarity matrices (RDMs) from source-projected MEG responses and modeled them as a combination of component RDMs using linear models. They then quantified the unique contribution of each component to a target RDM and found the inferior temporal cortex to contribute unique information to primary visual areas at latencies consistent with feedback. Finally, the authors trained different types of deep neural networks (DNNs) to mimic the representational dynamics of ventral visual areas and tested the resulting architectures on their ability to classify objects before and after successively deactivating (cooling) either lateral or top-down connections. They found that deactivation of top-down connections into the layer that most closely models V1-V3 had a strong effect on model performance, indicating a causal role of top-down connections in object classification.

Altogether, previous research has demonstrated a need to include non-feedforward information transfer into models of visual processing with more recent models generally attributing higher importance to feedback and top-down processes. The following paragraph focuses on how the structure of the visual world can be used to constrain theories about viable mechanisms for feedback.

Redundancy and natural image statistics

The human brain is subject to metabolic constraints. At the same time, rapid processing of vast amounts of visual input is often critical for appropriate behavioral responses to changes in the environment. Balancing the metabolic cost of neuronal functioning against the demands for speed, accuracy and reliability of perception requires that the brain rapidly and efficiently extracts relevant information from a stream of incoming data.

In the visual system this data is comprised of the changing patterns of light of various amplitudes and wavelengths projected onto the retinae. The information that needs to be extracted is e.g. whether and where an object is present in a cluttered visual scene. Importantly, the relationship between data and information is not necessarily a linear one, that is, more data does not always equal more information. Instead, signal redundancies such as repeating patterns account for a large portion of total data volume without contributing unique information.

The measure for the average rate of information transmitted by a signal is known as information entropy. Shannon (Shannon, 1948) famously formalized the relationship between units of information and the signal they are embedded in as

$$H(P) = -\sum_{i=1}^{N} p_i \log\,(p_i) \qquad (\mathbf{1})$$

Here, H(P) represents the (Shannon-) entropy of the random variable H with a distribution P, i.e. the amount of uncertainty in the signal. In the case of a single toss of a fair coin (1) simplifies to

$$\mathrm{H(P)} = -\sum_{\mathrm{i=1}}^{2} \frac{1}{2} \log_2 \frac{1}{2} \qquad (\mathbf{2})$$

$$= -\sum_{i=1}^{2} \frac{1}{2} \cdot (-1) = 1 \qquad (\mathbf{3})$$

This means that the outcome is equally likely to be either one of two states (heads or tail) and the amount of information transmitted with a single toss is one unit (1 bit). With lower uncertainty in the signal (i.e. an unfair coin) the amount of information generated by each toss is lower than 1. Analogous to tosses of coins, if one part of an image restricts the possible contents of another part, such as in a repeating, regular grating, the amount of entropy, and therewith information transmitted by said image, is lower than if the parts were independent of each other.

An important distinction needs to be made between a data streams' entropy and its signal to noise ratio (SNR). In a white noise image for example, the entropy is very high, because each pixel is independent of every other pixel and there are no redundancies. At the same time, white noise conveys no relevant information. Shannon's original formulation holds for the case of a noise-free channel. However, if data consists of some combination of signal and noise, as is the case in the electrical activity transmitted from the retina in response to a stimulus, focusing solely on the maximization of entropy/minimization of redundancy is counterproductive. Figure 4 illustrates the meaning of redundancy, entropy and SNR by revisiting the white noise image example: On a 5x5 pixel grid, the signal consists of a rhombus formed by fully black pixels, surrounded by fully white pixels (left box). The middle box contains pure noise with no discernable pattern and the rightmost box (c) contains a combination of box (a) and box (b), with a signal (box (a)) to noise (box (b)) ratio of 1, meaning that signal and noise are equally represented in the image. Between those three images, box (a) has the highest SNR, but also the highest amount of redundancy because of its pattern regularity. Box (b) has the highest entropy but contains no relevant signal and box (c) has higher entropy and reduced redundancy compared to box (a) but does not contain more (relevant) signal. The data displayed in (a) can be losslessly compressed, i.e. the information contained in it can be expressed with fewer information units than are necessary to code for each pixel individually, in a fully reversable process. The data in figure (b) cannot be reduced. The data in figure (c) can be reduced, but only via lossy compression, meaning that the process is not fully reversible. This is because the noise contained in (c) is uncorrelated and can itself not be compressed. Thus,

when considering redundancies in data within the visual system, it is important to take the presence of noise into account.

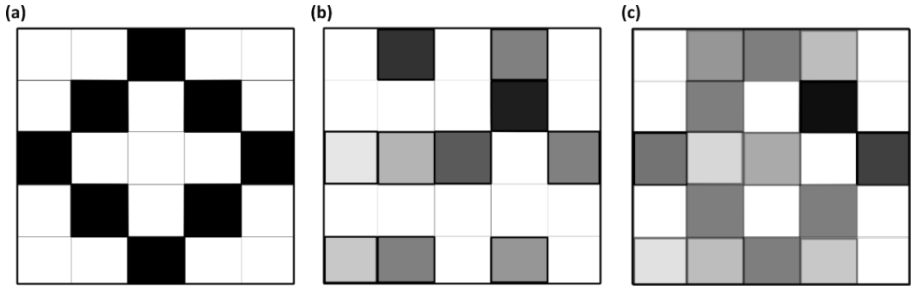


Figure 4: In the presence of noise, higher entropy does not equate to more information.

In natural viewing conditions, the human visual system encounters two main flavors of redundancy: temporal redundancy and spatial redundancy. Temporally redundant information corresponds to visual input that either remains stable across time or has a constant derivative across time (e.g. an object that moves with a steady speed in a straight line). Whether or not a stimulus is predictable from the recent history of sensory inputs has been demonstrated to have profound effects on visual adaptation (Aschner, Solomon, Landy, Heeger, & Kohn, 2018; Carandini, Barlow, O'keefe, Poirson, & Movshon, 1997). Contrary to laboratory settings where images often have an abrupt onset, last for the period of one to a few fixations, and then have an abrupt offset, natural vision is temporally contingent. From one fixation to the next, visual input remains largely stable and predictable. But although the saccadic interruptions of the continuous visual input alone form the basis of an entire research field (Dorr & Bex, 2013; Schütz, Braun, & Gegenfurtner, 2011), a detailed review of temporal elements in visual processing is beyond the scope of this thesis. Instead, the work presented here focuses on spatial redundancy within visual input.

Spatial redundancies correspond to statistical dependencies between neighboring locations in visual space as well as dependencies across spatial frequencies. As demonstrated by the success of common compression algorithms that reduce images to a fraction of their original size by exploiting the fact that neighboring pixels are often not independent from each other, the effective information content of images is considerably lower than what the system is initially confronted with. This is known as inter-pixel redundancy. To illustrate, the image in figure 1a,

a grayscale 640x480 image, occupies 307.2KB of storage as an uncompressed bitmap. After compression, storage size reduces to 110.6KB (LZW compression, lossless) or even 67.6KB (JPEG 17 format based on discrete cosine transform, lossy). Apart from the natural world equivalent to inter pixel redundancies, cross spatial frequency correlations can substantially decrease entropy in a visual scene. Figure 5 illustrates how low spatial frequency contrast is a good predictor of the location of high spatial frequency contrast.

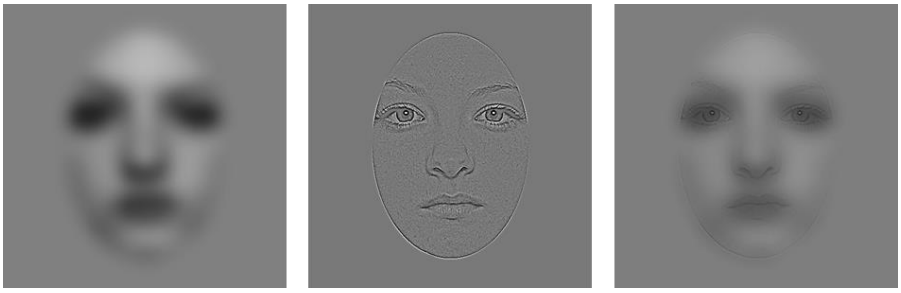


Figure 5: Low and high spatial frequencies (left and middle panels respectively) are highly correlated in natural images. The rightmost image is an overlay of the images on the left. The low spatial frequency contrast indexes where the high spatial frequency contrast is located

Apart from those low-level dependencies between spatial locations and frequencies within a single view, visual input is also characterized by intermediate and higher-level statistical dependencies.

A typical human's visual diet consists of a wide array of scenes, objects, faces, animals, drawings, text etc. But despite the diversity of visual input, on average, natural images share some statistical regularities. When decomposing a large set of natural images into their Fourier spectra, the average amplitude spectrum shows a roughly $1/f^{\alpha}$ slope (with $\alpha \sim= 1.2$ for e.g. natural scenes) with lower spatial frequencies contributing more to the total image contrast than higher spatial frequencies. Furthermore, especially in outdoor, natural scenes, cardinal (i.e. horizontal and vertical) orientations are predominant (Coppola, Purves, McCoy, & Purves, 1998; Torralba & Oliva, 2003). Within a given category of visual objects, the statistical commonalities they share are even more striking. Faces for example possess only marginally different Fourier-spectra between exemplars and contrast

is most pronounced in the vertical arrangement of horizontally elongated features such as the eyes or the mouth (Dakin & Watt, 2009).

Image recognition algorithms for computer vision applications, that are confronted with similar challenges of limited resources as the brain, routinely reduce temporal (Buckler, Bedoukian, Jayasuriya, & Sampson, 2018), spatial (Chen et al., 2019; Han et al., 2016; D. T. Nguyen, Zong, Ogunbona, & Li, 2010), coding (Welch, 1984; Ziv & Lempel, 1978), or semantic (Weinsberg et al., 2012) redundancies in the signal to alleviate computational load, and the emergence of computer-vision algorithms that are able to classify the content of visual scenes based on self-learned statistical models with increasingly high accuracy demonstrates, that low- to intermediate-level regularities are a good predictor of scene content.

However, even the most complex feature-based deep neural networks still consistently fail to recognize objects in the presence of noise, occlusion or even small levels of abstraction (Grm, Štruc, Artiges, Caron, & Ekenel, 2017; Kar, Kubilius, Schmidt, Issa, & DiCarlo, 2019; Moosavi-Dezfooli, Fawzi, & Frossard, 2016; A. Nguyen, Yosinski, & Clune, 2015) while human observers have less problems coping with a wide range of impediments. Thus, despite the substantial diagnostic value of intermediate level image statistics, they are not sufficient to resolve everyday visual tasks (but see also Neri, 2018). Correspondingly, humans have been shown to also use higher level semantic regularities, such as object – location binding (Notaro & Hasson, 2019), to process visual information.

The functional neuroanatomy of the human visual system is in principle suitable to detect and reduce image redundancies at different levels: In primary visual cortex, neighboring patches of cortex tend to respond to neighboring patches of the visual field (Tusa, Palmer, & Rosenquist, 1978). Thus, at the most fundamental level of visual perception, close range lateral inhibition between cells in the retina (Srinivasan, Laughlin, & Dubs, 1982) and primary visual cortex (Das & Gilbert, 1999) responding to close by, and therefore likely similar, input presents a simple and effective way to reduce close range image redundancies. The principle of lateral inhibition was proposed by Ernst Mach (1866) who discovered that the edges between parallel and adjacent bands of different shades of gray appear to have higher contrast than the same luminance difference further away from the boundary, a visual illusion known as Mach bands. Lateral inhibition in the visual system has since been described with respect to adjacent spatial frequency channels (Sagi & Hochstein, 1985), orientations (Blakemore, Carpenter, &

Georgeson, 1970) and motion detectors (Creutzfeldt, Kuhnt, & Benevento, 1974), but in its original formulation it was viewed as a mechanism of mutual suppression between neighboring cells or cells with a center surround receptive field (Ratliff, 1965). Although more recent accounts of luminance step illusions such as Mach bands propose an array of possible mechanisms ranging from local energy feature detectors (Concetta Morrone & Burr, 1988) to response normalization (Kingdom, 2014), the principal tenet that similar input is attenuated in favor of deviations from uniformity remains intact.

At the second level of abstraction from image space, spatial frequency redundancies that occur in natural images hold the potential of being exploited. It has been suggested that efficient processing of high spatial frequency information could be aided by low spatial frequency triggered feedback to early visual cortex (Goffaux et al., 2010; e.g. Kveraga, Boshyan, & Bar, 2007; Musel et al., 2014; Petras, ten Oever, Jacobs, & Goffaux, 2019). The underlying assumption is that low spatial frequency information is rapidly transmitted to cortex where it is quickly processed by extrastriate areas and generates feedback signals to facilitate the efficient extraction of high spatial frequency detail information. A possible mechanism to enable a temporal advantage of coarse information is given by the different conduction speeds of magnocellular and parvocellular pathways (Jean Bullier & Nowak, 1995; Petersen, Miezin, & Allman, 1988). M-cells are characterized by their rapid and transient response to achromatic, low resolution stimuli with a high sensitivity for luminance contrast (Merigan & Maunsell, 1993; Nowak & Bullier, 1997) whereas P-cells show later and more sustained responses, preferentially to chromatically defined, higher resolution stimuli and require considerably higher luminance contrast to respond to achromatically defined stimuli (Maunsell et al., 1999; Tootell, Hamilton, & Switkes, 1988). However, the contribution of M and P pathways to coarse-to-fine spatial frequency processing of naturalistic images has rarely been addressed directly. A notable exception is Kveraga et al. (2007) were the authors presented participants with line drawings of objects, manipulated to be M-biased (low luminance contrast, achromatic) or P-biased (isoluminant, red-green defined) while recording fMRI. Participants were faster and more accurate in judging the stimulus's real-life size when stimuli were M-biased. fMRI results revealed a greater activation of orbitofrontal cortex in response to M-biased compared to P-biased stimuli whereas the reverse pattern was found in fusiform and occipitotemporal regions. The authors interpret their findings as evidence for

M-cell driven feedback from orbitofrontal to ventral visual areas. In most studies however, rather than being optimized to selectively trigger M or P cells, the stimuliused to address the processing of visual scenes or complex visual objects are high contrast, grayscale images with a wide range of spatial frequencies adjusted to constrain contrast cycles per image or cycles per object as a proxy of coarseness rather than cycles per degree as a an absolute measure of retinal input (e.g. De Cesarei & Codispoti, 2013; Kauffmann, Chauvin, Guyader, & Peyrin, 2015; Oliva & Torralba, 2006). Nevertheless, evidence across absolute spatial scales and across diverse paradigms indicate a coarse to fine (Goffaux et al., 2010; Halit, de Haan, Schyns, & Johnson, 2006; Murray & Plainis, 2003; Vassilev & Mitov, 1976) and global to local (Navon, 1977) temporal processing precedence that could facilitate low spatial frequency guidance of high spatial frequency processing.

In the context of statistical regularities within visual categories, such as the stereotypical distribution of contrast in frontal views of faces, it may seem trivial that the visual information that is most prominent in the image is also contributing most to the observer's behavior. E.g. human observers strongly rely on low spatial frequency information when asked to detect faces in a stream of objects (Quek, Liu-Shuang, Goffaux, & Rossion, 2018) and on horizontal contrast when identifying individual faces (Dakin & Watt, 2009; Goffaux & Greenwood, 2016). When presented with naturalistic broadband scenes or outdoor scenes however, where the horizon contributes to a strong dominance of horizontal contrast, human observes show the greatest sensitivity to non-horizontal contrast, presumably because the horizon does not offer much useful information about the features of the scene (Essock, DeFord, Hansen, & Sinai, 2003; Hansen & Essock, 2004). Other studies have shown, that the internal models human observers employ to judge the orientation content of stimuli under noisy conditions, show biases matching the environmental statistics of orientation distribution (Girshick, Landy, & Simoncelli, 2011). This indicates that the visual system, rather than simply processing all the available distribution of contrast indiscriminately, can flexibly tune to the image characteristics that promise most informational value. Interestingly, a proposed biological implementation of such Bayesian-type inference based on image statistics operates via the basic principle of surround suppression that is flexibly gating neuronal activity based on the inferred amount of redundancy within a given stimulus (Coen-Cagli, Kohn, & Schwartz, 2015).

At the highest level of abstraction from primary visual input, semantic context shapes stimulus expectation (Bar, 2004; Kaiser, Quek, Cichy, & Peelen, 2019). E.g. a car is more expected, and therefore more likely to be detected, in the context of a street, compared to a forest scene. Brandman & Peelen (2017) demonstrated that the multivariate representation, i.e. the spatial pattern of voxel activations specific to a visual object, of a degraded object measured with fMRI was dependent of the context in which it was embedded. They further showed that the facilitation of object representations by the surrounding scene as expressed in MEG decoding scores peaked at a latency consistent with high-level feedback to visual cortex. These findings indicate that implicit knowledge about the abstract statistical regularities of the environment can be employed by the visual system to support perception.

In summary, while redundancy is computationally and metabolically disadvantageous when considering pure information transfer, it can be exploited to reduce resources needed for information processing. The visual world contains copious redundancies on different levels. While low-level redundancies such as inter-pixel- and spatial frequency redundancy can potentially be resolved via lateral inhibition by a naïve system with no knowledge of statistical regularities, intermediate level natural image statistics as well as high-level contextual cues require a more complex interplay between different visual, and perhaps even semantic subsystems. Human vision is supported by a dynamic relationship between feedforward, recurrent, and feedback connections, suitable to exploit all levels of redundancy to reduce computational load.

Efficient coding

When talking about the reduction of computational load it is essential not to overextend the analogy of a computer. While conceptualizing visual perception in computational terms has contributed greatly to the development of modern vision theory and still provides a basis to characterize the steps and transformations necessary to derive semantic meaning from patterns of light, its premise is limited. If the brain would simply reduce redundancy the same way as most compression algorithms to preserve computational resources, we should expect a progressive reduction of redundancy as well as resources dedicated to the processing of visual stimuli at each hierarchical level.

But contrary to a computer, the brain is evolved rather than designed and empirical evidence seems to support an increase, rather than a decrease of redundancy along the visual pathway (Wandell, 1995).

The most efficient way to non-redundantly code information would be a distributed representation in which neurons fire with high and close to equal frequencies (Field, 1994). Instead, empirical evidence suggests that rather than forming a sparse distribution, ensembles of neurons with similar tuning properties cluster together and firing rates are considerably lower than optimal efficiency would demand (Olshausen & Field, 1997). Responses of individual neurons in close proximity to each other are very similar and a considerable amount of information can be retrieved from summarizing the redundancies within those patches as evidenced by the success of methods like EEG and MEG, both of which rely on the synchronous activity of thousands of neurons to measure minute changes in polarization (Nunez & Harth, 1982). Further, the number of cells increases from subcortical to cortical stages of visual processing. E.g. the amount of retinal ganglion cells (approx. $2x10^6$) is considerably lower than the number of V1 cells (approx. 10^9) (Barlow, 2001; Braitenberg & Schüz, 2013). On top of the sheer difference in cell count, the channel capacity, i.e. the amount of information that can pass through a given channel (in the brain largely determined by e.g. the mean firing rate and the dynamic range of a neuron) increases. Retinal ganglion cells for example receive their input from comparatively slow photoreceptors that hardly ever operate within their full dynamic range while V1 cells show generally higher firing rates. Taken together, the clustering of cells with similar response profiles, as well as the increase in cell number and firing rates supports an increase in redundancy and computational resources devoted to the processing of the visual signal.

To reconcile the apparent increase in computationally and metabolically expensive redundant representations employed by the brain with the selective advantage required to favor the evolution of such a system, Barlow (2001) suggests to treat environmental redundancies as a source of information on their own rather than an obstacle to efficient coding. He states that in an inherently noisy system like the brain, scattered neurons firing at high and similar frequencies are statistically disadvantageous. This is because such (redundancy reducing) high activity-ratio, distributed representations would lead to inaccuracies when estimating probabilities from the frequency at which events occur. In order to make sense of the statistical patterns (which per definition contain redundancies) it encounters,

the brain must have a way to count how often a given event, such as a type of visual stimulus, occurs and under which circumstances it does so.

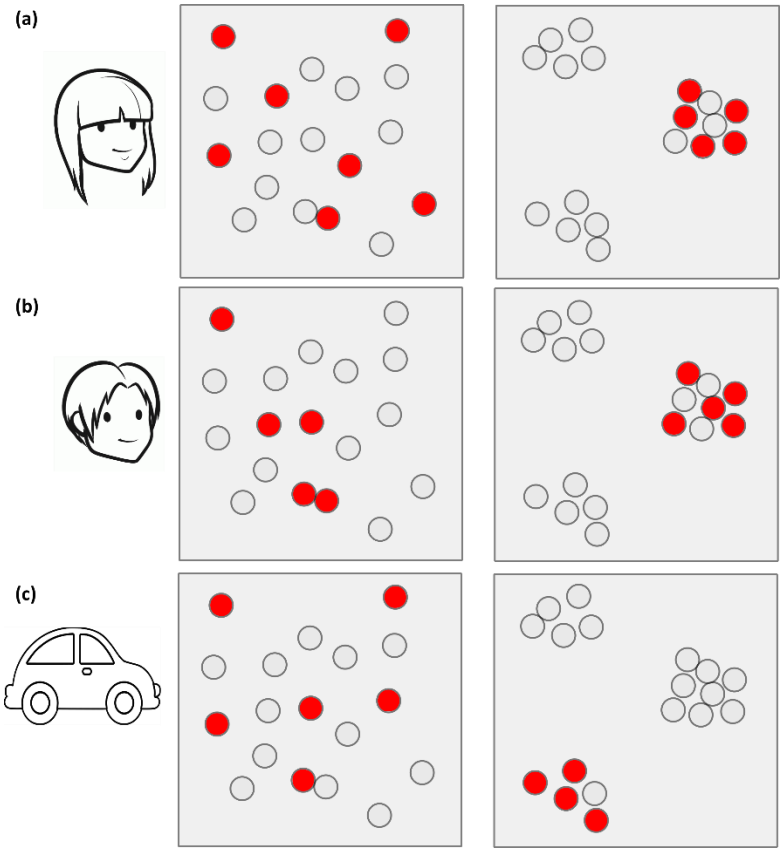


Figure 6: In a distributed representation (left column), cells are likely to be active during more than one condition whereas modular organizations (right column) reduce the overlap between representations. To count the number of occurrences of an event, e.g. the presence of a face, requires only the counting of elements that are uniquely active whenever there is a face if the representation is modular. In the distributed case, individual cells are active in many different representations and thus not in themselves a reliable indicator of any given class of events.

For example, to build a size invariant representation of a face, the brain must learn second order relations between facial features such as the common relative

distances between sets of eyes and mouths (see Fiser & Aslin, 2001 for statistical learning of higher order structure and Rotshtein, Geng, Driver, & Dolan, 2007 for second order relations in face processing). In a distributed representation, the counting of occurrences of one event type, e.g. faces, would face interference from the occurrence of different event types, e.g. cars (see figure 6). To disambiguate visual input by adding extra cells to track and count event occurrences, a distributed, input redundancy efficient system would require more cells than are found in the whole brain.

Instead, Gardner-Medwin & Barlow (2001) argue that in order to increase statistical efficiency (see Fisher, 1925 and Pelli, Barlow, 1987) a modular organization of direct representations would be required. To simultaneously preserve the flexibility to detect unforeseen or rare events, a distributed representation with limited representational overlap (low number of representational elements active during several possible events) and high levels of redundancy is optimal. In such a system, signal redundancies such as co-occurrence of contrast across spatial scales, could be used to increase certainty under noisy stimulus conditions. As illustrated in figure 5, the low spatial frequency image contrast acts as an index to flag the location of high spatial frequency information. Thus, information *about* those redundancies could be beneficial to restrict the processing of high spatial frequency detail to relevant locations in visual space, therewith ultimately reducing computational demand.

In the human visual system, regions within the lateral occipital and inferior temporal cortex that are most commonly associated with the processing of specific object categories such as the occipital face area (OFA) (Pitcher, Walsh, & Duchaine, 2011; Rossion, Caldara, et al., 2003; Rossion, Schiltz, & Crommelinck, 2003), the fusiform face area (FFA) (Kanwisher & Yovel, 2006) or the parahippocampal place area (PPA) (Epstein, Harris, Stanley, & Kanwisher, 1999) have been shown to prefer stimuli that have similar statistical properties as their preferred object category (Andrews, Clarke, Pell, & Hartley, 2010; Rossion, Hanseeuw, & Dricot, 2012). Such neuronal populations that respond to both basic image statistics and higher-level semantic image content would be ideally suited to provide early visual cortex with information about redundancies in spatial frequency or orientation content along with the information about which spatial locations are most relevant for higher resolution processing.

In summary, the visual system employs feedback connections to make sense of noisy visual input and extract information from vast amounts of data. Visual stimuli naturally contain high levels of redundancy and statistical regularities. In the interest of accurate and statistically efficient representations, it would be most beneficial for the brain to make those redundancies explicit and use them rather than discard them. A computationally and biologically feasible way of doing so is to engage feedback connections from higher level to primary visual cortex.

The common theme of the work presented in this thesis is the question how the brain uses image redundancies to facilitate the efficient processing of complex visual stimuli. The global conjectures that determined the experimental design and the more specific questions addressed in the empirical chapters were, in coarse-to-fine order:

- The visual system is sensitive to statistical regularities in the environment.
- Visual areas with complex receptive fields, such as those found in the ventral visual pathway, influence early visual cortex processing via feedback connections.
- Coarse image information influences the processing of correlated image details.
- This influence is largely suppressive, i.e. the more information coarse image structure contains about image detail, the lower the resources dedicated to detail-processing.

As all empirical chapters presented here strongly rely on image processing to produce the stimuli used for experimentation, the following gives a brief overview of the basic principles of the 2-dimensional Fourier transform that forms the basis of most image manipulations used in the presented studies. A more critical evaluation of the potential pitfalls in image processing is given in the general discussion.

Basics of the 2-dimensional Fourier transform

Every image can be broken down into a set of 2D sinusoids of different frequencies, orientations, and phases. As equation (1) shows, the discrete 2D Fourier transform (dFFT2) achieves this by first transforming the pixels along the columns in a matrix representation of the image and then transforming the rows of the resulting matrix according to

$$(1) \qquad F(u,v) = \frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y)exp\left[-i2\pi\left(\frac{ux}{M}+\frac{vy}{N}\right)\right] \quad for \begin{smallmatrix} u=0,1,2,\ldots,M-1 \\ v=0,1,2,\ldots,N-1 \end{smallmatrix}$$

Where M and N are the numbers of samples (i.e. pixels) in the rows and columns of the image matrix respectively and *i* is the imaginary unit. F(u,v) is the 2D complex spectrum of the image f(x,y). The term

$$(2) \quad \sum_{j=0}^{n-1} exp\left[-\frac{i2\pi}{n_k}\right]$$

is the familiar 1D discrete Fourier transform. This goes to show that the 2D (or n-D) Fourier transform is equivalent to calculating a 1-dimensional FFT for each dimension of the input matrix and all properties of the 1D FFT, such as the uncertainty principle (events localized in time produce a spread in the frequency spectrum and vice versa) or stationarity constraints apply equally to n-D FFTs.
The result of dFFT2 is an array of complex values that can be decomposed into an amplitude and a phase spectrum with the lowest frequencies of the amplitude spectrum represented in the outermost corners of the array. To aide visualization and further processing, the amplitude spectrum is commonly rearranged such that the lowest frequency value, corresponding to the mean luminance of the image, forms the center of the matrix and the surrounding values correspond to frequencies increasing with eccentricity. Horizontal orientations in the FFT correspond to vertical orientations in the image and vice versa. This rearrangement is known as FFT shifting. Because the dFFT2 is an invertible, linear transformation, the original image can be recovered fully from it. This property permits to manipulate images in the Fourier domain, where features like spatial frequency and orientation are readily accessible. Commonly used algorithms alter spatial

frequency and orientation spectra (i.e. filtering) or manipulate phase information (e.g. phase scrambling). However, some constraints need to be considered when working with digital images that have a finite resolution and dynamic range, i.e. the number of resolvable steps between the lowest and the highest luminance value.
The highest spatial frequency that can be represented in an image is half the sampling frequency, i.e. the image size in pixels, also known as the Nyquist frequency (Grenander, 1959). In practice, this means alternating between two pixel values every pixel. Any spatial frequency higher than the Nyquist frequency will be undersampled and lead to aliasing (see figure 7 a3 and b). Because of the limited number of gray levels a regular computer monitor can resolve, usually 256, there is also a lower bound to the spatial frequencies that can be represented as a smoothly varying sine wave. Image 6b2 shows an example of an image created as a 256x256 pixel grating that smoothly varies from black to white in 255 steps (note that scaling errors due to a mismatch between the print resolution and the image resolution apply). This corresponds to a spatial frequency of 1cycle over 512 pixels (only half a cycle is shown here, see superimposed blue trace). Any spatial frequency lower than this would have to be approximated by repeating pixel values. Similarly, the minimum luminance step, i.e. the minimum non-zero difference between two pixels is given by 1/256. The limited number of grayscale values becomes particularly relevant when producing complex gratings that consist of many different sine waves with different phases and amplitudes, such as images of natural scenes or objects where the local dynamic range can differ vastly across different image regions. Figure 7 illustrates some basic properties of Fourier transforms and gives examples of the dFFT2 of gratings and simple shapes. A more thorough discussion of the digital image processing choices made in the empirical chapters of this thesis follows in chapter 6 (General Discussion).

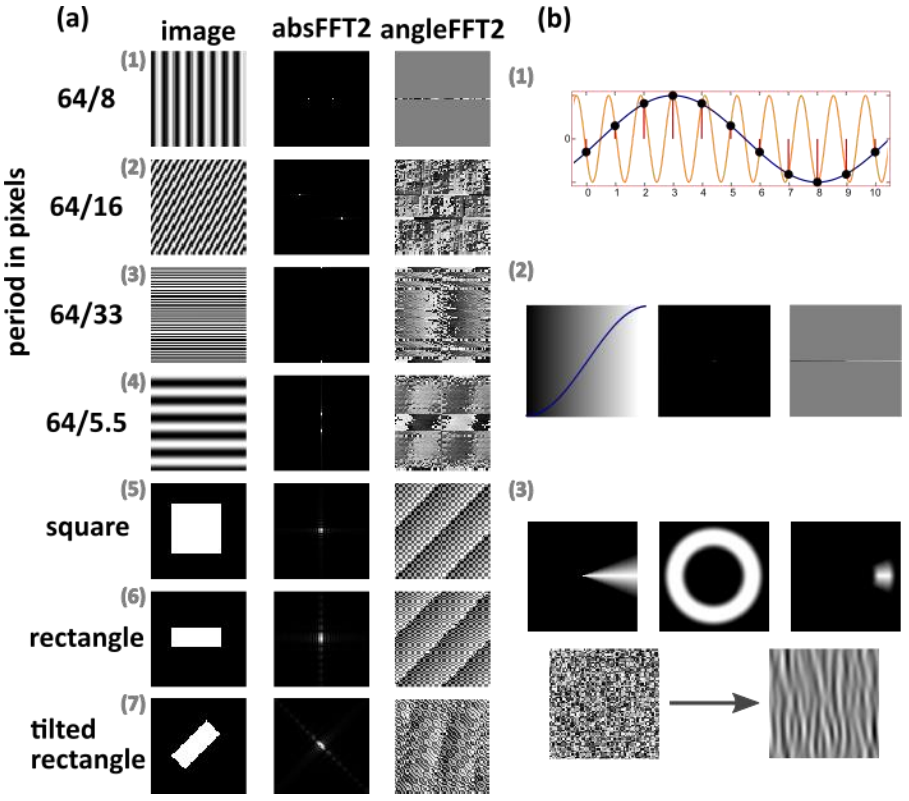


Figure 7: (a) Examples of 64x64 pixel 2D sinewave gratings and simple shapes along with their corresponding dFFT2 amplitude and phase spectra. The shortest period that can be unambiguously represented in 64 pixels is 64/32 (i.e. two pixels) long and would correspond to alternating between two pixel-values (e.g. 1010…10). When an image contains an integer number of cycles of a sine wave grating, such as in (a)(1) and (a)(2), the shifted amplitude spectrum shows two isolated >0 pixels, equidistant from the center, corresponding to power at a single frequency along the orientation of the grating. (a)(3) shows a grating with a sine wave frequency above the Nyquist frequency for this image. When the grating's frequency exceeds this limit, more than one frequency fits the sampling points (see 1b) leading to the spurious representation of the alternative frequency (i.e. aliasing). 1(a)(4) shows the result of plotting a non-integer number of cycles of a sine wave. Although only a single frequency sinewave constitutes the image, the Fourier spectrum shows some spurious non-zero amplitudes around the two expected amplitude peaks. 1(a)(5)-1(a)(7) show simple rectangular shapes. No single sine wave can describe the

combination of sharp edges and a large uniform surface, leading to a spread in the Fourier spectrum. When the object is rotated by 45° the edges of the pixels do not line up with the edge of the figure, leading to a serrated contour and large spread in the Fourier spectrum. Phase spectra are more difficult to interpret than amplitude spectra. It is however noteworthy, that they are closely related to the amplitude spectra of images, although in filtering or phase scrambling procedures they are often treated as independent (see discussion on phase scrambling in chapter 6). 6(b)(1) demonstrates that if the sampling rate is below twice the highest frequency in the signal, the sampled points can be equally well fitted by more than one sinewave. Consequently, the Fourier amplitude spectrum has non-zero values at more than one frequency as seen in 6(a)(3). 6(b)(2) shows a smooth luminance increase from zero to 256 in 255 steps along with its amplitude and phase spectra. 6(b)(3) shows a spatial frequency and orientation filter. The wedge restricts edge orientations to a range around vertical with attenuation linearly increasing with angular distance from the center orientation. The annulus restricts spatial frequency to a range around the center spatial frequency with a cosine tapered increase in attenuation away from the center frequency both towards and away from the center of the FFT. Below is an example white noise image before and after multiplying its FFT amplitude spectrum with the product of the wedge and the annulus filter. Note that all images in (a) have been produced as 64x64 pixel grids and all images in (b) as 256x256 pixel grids. However, scaling errors apply to the figure.

Outline of Thesis

Most empirical chapters in some way address the integration of different ranges of spatial frequencies while considering their relationship in image space. In chapter two we map out the spatial frequency sensitivity profiles of different regions of interest from the primary visual cortex along the inferior temporal lobes up to orbitofrontal regions. We compare between responses to simple sine wave gratings and more ecologically relevant 1/f noise patches. Chapter three then directly addresses the integration of different ranges of spatial frequency in the processing of images of human faces and monkey faces as well as their phase scrambled counterparts. We demonstrate that relevant low spatial frequency information facilitates the integration of high spatial frequency detail. Chapter four follows up on the findings of chapter three and extends them by investigating the origin and

information content of feedback to early visual cortex as a modulating force for spatial frequency integration. Chapter five aims to dissociate between feedback originating in mid-level visual regions V2-V4 and higher level inferior temporal regions by selectively silencing either object recognition alone or object recognition and basic image shape and phase coherence together. Finally, chapter six contains a general discussion of the topics raised in chapters two to five.

References

Ahissar, M., & Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature*, *387*(6631), 401–406. https://doi.org/10.1038/387401a0

Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Sciences*, *8*(10), 457–464.

Andrews, T. J., Clarke, A., Pell, P., & Hartley, T. (2010). Selectivity for low-level features of objects in the human ventral stream. *Neuroimage*, *49*(1), 703–711.

Aschner, A., Solomon, S. G., Landy, M. S., Heeger, D. J., & Kohn, A. (2018). Temporal contingencies determine whether adaptation strengthens or weakens normalization. *Journal of Neuroscience*, *38*(47), 10129–10142.

Bar, M. (2004). Visual objects in context. *Nature Reviews. Neuroscience*, *5*(8), 617.

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., … Rosen, B. R. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(2), 449–454.

Barlow, H. (2001). Redundancy reduction revisited. *Network: Computation in Neural Systems*, *12*(3), 241–253.

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.

Beck, D. M., & Kastner, S. (2009). Top-down and bottom-up mechanisms in biasing competition in the human brain. *Vision Research*, *49*(10), 1154–1165.

Blakemore, C., Carpenter, R. H. S., & Georgeson, M. A. (1970). Lateral inhibition between orientation detectors in the human visual system. *Nature*, *228*(5266), 37–39.

Bouvier, S., & Treisman, A. (2010). Visual feature binding requires reentry. *Psychological Science*, *21*(2), 200–204.

Braitenberg, V., & Schüz, A. (2013). *Anatomy of the cortex: statistics and geometry* (Vol. 18). Springer Science & Business Media.

Brandman, T., & Peelen, M. V. (2017). Interaction between scene and object processing revealed by human fMRI and MEG decoding. *Journal of Neuroscience*, *37*(32), 7700–7710.

Buckler, M., Bedoukian, P., Jayasuriya, S., & Sampson, A. (2018). EVA^2: Exploiting Temporal Redundancy in Live Computer Vision. In *2018 ACM/IEEE 45th Annual International Symposium on Computer Architecture (ISCA)* (pp. 533–546). IEEE.

Bullier, J, McCourt, M. E., & Henry, G. H. (1988). Physiological studies on the feedback connection to the striate cortex from cortical areas 18 and 19 of the cat. *Experimental Brain Research*, *70*(1), 90–98.

Bullier, Jean. (2001). Integrated model of visual processing. *Brain Research Reviews*, *36*(2), 96–107.

Bullier, Jean, & Nowak, L. G. (1995). Parallel versus serial processing: new vistas on the distributed organization of the visual system. *Current Opinion in Neurobiology*, *5*(4), 497–503.

Burkhalter, A., & Bernardo, K. L. (1989). Organization of corticocortical connections in human visual cortex. *Proceedings of the National Academy of Sciences*, *86*(3), 1071–1075.

Carandini, M., Barlow, H. B., O'keefe, L. P., Poirson, A. B., & Movshon, J. A. (1997). Adaptation to contingencies in macaque primary visual cortex. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *352*(1358), 1149–1154.

Chen, Y., Fang, H., Xu, B., Yan, Z., Kalantidis, Y., Rohrbach, M., … Feng, J. (2019). Drop an octave: Reducing spatial redundancy in convolutional neural networks with octave convolution. *ArXiv Preprint ArXiv:1904.05049*.

Coen-Cagli, R., Kohn, A., & Schwartz, O. (2015). Flexible gating of contextual influences in natural vision. *Nature Neuroscience*, *18*(11), 1648.

Concetta Morrone, M., & Burr, D. C. (1988). Feature detection in human vision: A phase-dependent energy model. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *235*(1280), 221–245.

Coogan, T. A., & Burkhalter, A. (1990). Conserved patterns of cortico-cortical connections define areal hierarchy in rat visual cortex. *Experimental Brain Research*, *80*(1), 49–53.

Coppola, D. M., Purves, H. R., McCoy, A. N., & Purves, D. (1998). The distribution of oriented contours in the real world. *Proceedings of the National Academy of Sciences*, *95*(7), 4002–4006.

Creutzfeldt, O. D., Kuhnt, U., & Benevento, L. A. (1974). An intracellular analysis of visual cortical neurones to moving stimuli: responses in a co-operative neuronal network. *Experimental Brain Research*, *21*(3), 251–274.

Dakin, S. C., & Watt, R. J. (2009). Biological "bar codes" in human faces. *Journal of Vision*, *9*(4), 2.

Das, A., & Gilbert, C. D. (1999). Topography of contextual modulations mediated by short-range interactions in primary visual cortex. *Nature*, *399*(6737), 655.

De Cesarei, A., & Codispoti, M. (2013). Spatial frequencies and emotional perception. *Reviews in the Neurosciences*, *24*(1), 89–104.

Dorr, M., & Bex, P. J. (2013). Peri-saccadic natural vision. *Journal of Neuroscience*, *33*(3), 1211–1217.

Edelman, G. M. (1978). Group selection and phasic re-entrant signalling: a theory of higher brain function. *The Mindful Brain*, 55–100.

Edelman, G. M. (1993). Neural Darwinism: selection and reentrant signaling in higher brain function. *Neuron*, *10*(2), 115–125.

Epstein, R., Harris, A., Stanley, D., & Kanwisher, N. (1999). The parahippocampal place area: recognition, navigation, or encoding? *Neuron*, *23*(1), 115–125.

Essock, E. A., DeFord, J. K., Hansen, B. C., & Sinai, M. J. (2003). Oblique stimuli are seen best (not worst!) in naturalistic broad-band stimuli: A horizontal effect. *Vision Research*, *43*(12), 1329–1335.

Field, D. J. (1994). What is the goal of sensory coding? *Neural Computation*, *6*(4), 559–601.

Firestone, C., & Scholl, B. J. (2016). Cognition does not affect perception: Evaluating the evidence for "top-down" effects. *Behavioral and Brain Sciences*, *39*.

Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, *12*(6), 499–504.

Fisher, R. A. (1925). Theory of statistical estimation. In *Mathematical Proceedings of the Cambridge Philosophical Society* (Vol. 22, pp. 700–725). Cambridge University Press.

Gardner-Medwin, A. R., & Barlow, H. B. (2001). The limits of counting accuracy in distributed neural representations. *Neural Computation*, *13*(3), 477–504.

Girshick, A. R., Landy, M. S., & Simoncelli, E. P. (2011). Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. *Nature Neuroscience*, *14*(7), 926.

Goffaux, V., & Greenwood, J. A. (2016). The orientation selectivity of face identification. *Scientific Reports*, *6*, 34204.

Goffaux, V., Peters, J., Haubrechts, J., Schiltz, C., Jansma, B., & Goebel, R. (2010). From coarse to fine? Spatial and temporal dynamics of cortical face processing. *Cerebral Cortex*, *21*(2), 467–476.

Gray, C. M., Engel, A. K., König, P., & Singer, W. (1990). Stimulus-dependent neuronal oscillations in cat visual cortex: Receptive field properties and feature dependence. *European Journal of Neuroscience*, *2*(7), 607–619.

Grenander, U. (1959). The Nyquist Frequency is that frequency whose period is two sampling intervals. *Probability and Statistics: The Harald Cramér Volume*, 434.

Grm, K., Štruc, V., Artiges, A., Caron, M., & Ekenel, H. K. (2017). Strengths and weaknesses of deep learning models for face recognition against image degradations. *IET Biometrics*, *7*(1), 81–89.

Halit, H., de Haan, M., Schyns, P. G., & Johnson, M. H. (2006). Is high-spatial frequency information used in the early stages of face detection? *Brain Research*, *1117*(1), 154–161. https://doi.org/10.1016/j.brainres.2006.07.059

Han, S., Pool, J., Narang, S., Mao, H., Gong, E., Tang, S., … Tran, J. (2016). DSD: Dense-sparse-dense training for deep neural networks. *ArXiv Preprint ArXiv:1607.04381*.

Hansen, B. C., & Essock, E. A. (2004). A horizontal bias in human visual processing

of orientation and its correspondence to the structural components of natural scenes. *Journal of Vision*, *4*(12), 5.

Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron*, *36*(5), 791–804.

Hubel, D. H., & Wiesel, T. N. (1977). Ferrier lecture-Functional architecture of macaque monkey visual cortex. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *198*(1130), 1–59.

Kaiser, D., Quek, G. L., Cichy, R. M., & Peelen, M. V. (2019). Object vision in a structured world. *Trends in Cognitive Sciences*.

Kanwisher, N., & Yovel, G. (2006). The fusiform face area: a cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *361*(1476), 2109–2128.

Kar, K., Kubilius, J., Schmidt, K., Issa, E. B., & DiCarlo, J. J. (2019). Evidence that recurrent circuits are critical to the ventral stream's execution of core object recognition behavior. *Nature Neuroscience*, *22*(6), 974–983.

Kauffmann, L., Chauvin, A., Guyader, N., & Peyrin, C. (2015). Rapid scene categorization: Role of spatial frequency order, accumulation mode and luminance contrast. *Vision Research*, *107*, 49–57.

Kennedy, H., & Bullier, J. (1985). A double-labeling investigation of the afferent connectivity to cortical areas V1 and V2 of the macaque monkey. *Journal of Neuroscience*, *5*(10), 2815–2830.

Kietzmann, T. C., Spoerer, C. J., Sörensen, L. K. A., Cichy, R. M., Hauk, O., & Kriegeskorte, N. (2019). Recurrence is required to capture the representational dynamics of the human visual system. *Proceedings of the National Academy of Sciences*, *116*(43), 21854–21863.

Kingdom, F. A. A. (2014). Mach bands explained by response normalization. *Frontiers in Human Neuroscience*, *8*, 843.

Kok, P., Failing, M. F., & de Lange, F. P. (2014). Prior expectations evoke stimulus templates in the primary visual cortex. *Journal of Cognitive Neuroscience*, *26*(7), 1546–1554.

Kok, P., Jehee, J. F. M., & De Lange, F. P. (2012). Less is more: expectation sharpens representations in the primary visual cortex. *Neuron*, *75*(2), 265–270.

Kveraga, K., Boshyan, J., & Bar, M. (2007). Magnocellular projections as the trigger of top-down facilitation in recognition. *Journal of Neuroscience*, *27*(48), 13232–13240.

Lamme, V. A. (1995). The neurophysiology of figure-ground segregation in primary visual cortex. *Journal of Neuroscience*, *15*(2), 1605–1615.

Lee, T. S., Mumford, D., Romero, R., & Lamme, V. A. F. (1998). The role of the primary visual cortex in higher level vision. *Vision Research*, *38*(15–16), 2429–2454.

Lennie, P. (1998). Single units and visual cortical organization. *Perception*, *27*(8), 889–935.

Liang, H., Gong, X., Chen, M., Yan, Y., Li, W., & Gilbert, C. D. (2017). Interactions between feedback and lateral connections in the primary visual cortex. *Proceedings of the National Academy of Sciences*, *114*(32), 8637–8642.

Logothetis, N. K., & Sheinberg, D. L. (1996). Visual object recognition. *Annual Review of Neuroscience*, *19*(1), 577–621.

Mach, E. (1866). *Über die physiologische Wirkung räumlich vertheilter Lichtreize: dritte Abhandlung: vorgelegt in der Sitzung vom 4. October 1866*. K. k. Hof- und Staatsdruckerei.

Markov, N. T., Vezoli, J., Chameau, P., Falchier, A., Quilodran, R., Huissoud, C., … Ullman, S. (2014). Anatomy of hierarchy: feedforward and feedback pathways in macaque visual cortex. *Journal of Comparative Neurology*, *522*(1), 225–259.

Marr, D. (1982). Vision: A computational investigation into the human representation and processing of visual information. *WH San Francisco: Freeman and Company*, *1*(2).

Maunsell, J. H. R., Ghose, G. M., Assad, J. A., McADAMS, C. J., Boudreau, C. E., & Noerager, B. D. (1999). Visual response latencies of magnocellular and parvocellular LGN neurons in macaque monkeys. *Visual Neuroscience*, *16*(1), 1–14.

Maunsell, J. H. R., & Newsome, W. T. (1987). Visual processing in monkey extrastriate cortex. *Annual Review of Neuroscience*, *10*(1), 363–401.

Merigan, W. H., & Maunsell, J. H. R. (1993). How parallel are the primate visual pathways? *Annual Review of Neuroscience*, *16*(1), 369–402.

Michalareas, G., Vezoli, J., Van Pelt, S., Schoffelen, J.-M., Kennedy, H., & Fries, P. (2016). Alpha-beta and gamma rhythms subserve feedback and feedforward influences among human visual cortical areas. *Neuron*, *89*(2), 384–397.

Milner, A. D., & Goodale, M. A. (1995). *Oxford: Oxford UP; 1995*. *The visual brain in action*.

Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: two cortical pathways. *Trends in Neurosciences*, *6*, 414–417.

Moosavi-Dezfooli, S.-M., Fawzi, A., & Frossard, P. (2016). Deepfool: a simple and accurate method to fool deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2574–2582).

Muckli, L., De Martino, F., Vizioli, L., Petro, L. S., Smith, F. W., Ugurbil, K., … Yacoub, E. (2015). Contextual feedback to superficial layers of V1. *Current Biology*, *25*(20), 2690–2695.

Murray, I. J., & Plainis, S. (2003). Contrast coding and magno/parvo segregation revealed in reaction time studies. *Vision Research*, *43*(25), 2707–2719.

Musel, B., Kauffmann, L., Ramanoël, S., Giavarini, C., Guyader, N., Chauvin, A., &

Peyrin, C. (2014). Coarse-to-fine categorization of visual scenes in scene-selective cortex. *Journal of Cognitive Neuroscience*, *26*(10), 2287–2297.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Perception and Psychophysics*, *5*, 197–200.

Neri, P. (2018). Classification images as descriptive statistics. *Journal of Mathematical Psychology*, *82*, 26–37.

Neumann, H., & Sepp, W. (1999). Recurrent V1–V2 interaction in early visual boundary processing. *Biological Cybernetics*, *81*(5–6), 425–444.

Nguyen, A., Yosinski, J., & Clune, J. (2015). Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 427–436).

Nguyen, D. T., Zong, Z., Ogunbona, P., & Li, W. (2010). Object detection using non-redundant local binary patterns. In *2010 IEEE International Conference on Image Processing* (pp. 4609–4612). IEEE.

Notaro, G., & Hasson, U. (2019). Spatial and semantic regularities produce interactive effects in early stages of visual orientation. *BioRxiv*, 639385. https://doi.org/10.1101/639385

Nowak, L. G., & Bullier, J. (1997). The timing of information transfer in the visual system. In *Extrastriate cortex in primates* (pp. 205–241). Springer.

Nunez, P. L., & Harth, E. (1982). Electric Fields of the Brain: The neurophysics of EEG. *Physics Today*, *35*, 59.

Oliva, A., & Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. *Progress in Brain Research*, *155*, 23–36.

Olshausen, B. A., & Field, D. J. (1997). Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Research*, *37*(23), 3311–3325.

Pelli, D. G., & Barlow, H. B. (1987). *Statistical Efficiency of Natural and Artificial Vision*. Optical Society of America. Retrieved from https://books.google.nl/books?id=EymgZwEACAAJ

Petersen, S. E., Miezin, P. M., & Allman, J. M. (1988). Transient and sustained responses in four extrastriate visual areas of the owl monkey. *Experimental Brain Research*, *70*(1), 55–60.

Petras, K., ten Oever, S., Jacobs, C., & Goffaux, V. (2019). Coarse-to-fine information integration in human vision. *NeuroImage*, *186*, 103–112.

Petro, L. S., & Muckli, L. (2017). The laminar integration of sensory inputs with feedback signals in human cortex. *Brain and Cognition*, *112*, 54–57.

Pitcher, D., Walsh, V., & Duchaine, B. (2011). The role of the occipital face area in the cortical face perception network. *Experimental Brain Research*, *209*(4), 481–493.

Quek, G. L., Liu-Shuang, J., Goffaux, V., & Rossion, B. (2018). Ultra-coarse, single-glance human face detection in a dynamic visual stream. *NeuroImage*.

Ratliff, F. (1965). *Mach bands: quantitative studies on neural networks* (Vol. 2).

Holden-Day, San Francisco London Amsterdam.
Rockland, K. S., & Virga, A. (1989). Terminal arbors of individual “Feedback” axons projecting from area V2 to V1 in the macaque monkey: a study using immunohistochemistry of anterogradely transported Phaseolus vulgaris-leucoagglutinin. *Journal of Comparative Neurology*, *285*(1), 54–72.
Roelfsema, P. R., Lamme, V. A. F., Spekreijse, H., & Bosch, H. (2002). Figure—ground segregation in a recurrent network architecture. *Journal of Cognitive Neuroscience*, *14*(4), 525–537.
Rossion, B., Caldara, R., Seghier, M., Schuller, A., Lazeyras, F., & Mayer, E. (2003). A network of occipito-temporal face-sensitive areas besides the right middle fusiform gyrus is necessary for normal face processing. *Brain*, *126*(11), 2381–2395.
Rossion, B., Hanseeuw, B., & Dricot, L. (2012). Defining face perception areas in the human brain: a large-scale factorial fMRI face localizer analysis. *Brain and Cognition*, *79*(2), 138–157.
Rossion, B., Schiltz, C., & Crommelinck, M. (2003). The functionally defined right occipital and fusiform “face areas” discriminate novel from visually familiar faces. *Neuroimage*, *19*(3), 877–883.
Rotshtein, P., Geng, J. J., Driver, J., & Dolan, R. J. (2007). Role of features and second-order spatial relations in face discrimination, face recognition, and individual face skills: Behavioral and functional magnetic resonance imaging data. *Journal of Cognitive Neuroscience*, *19*(9), 1435–1452.
Sagi, D., & Hochstein, S. (1985). Lateral inhibition between spatially adjacent spatial-frequency channels? *Perception & Psychophysics*, *37*(4), 315–322.
Schütz, A. C., Braun, D. I., & Gegenfurtner, K. R. (2011). Eye movements and perception: A selective review. *Journal of Vision*, *11*(5), 9.
Self, M. W., Kooijmans, R. N., Supèr, H., Lamme, V. A., & Roelfsema, P. R. (2012). Different glutamate receptors convey feedforward and recurrent processing in macaque V1. *Proceedings of the National Academy of Sciences*, *109*(27), 11031–11036.
Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*(3), 379–423.
Smith, A. T., Singh, K. D., Williams, A. L., & Greenlee, M. W. (2001). Estimating receptive field size from fMRI data in human striate and extrastriate visual cortex. *Cerebral Cortex*, *11*(12), 1182–1190.
Srinivasan, M. V., Laughlin, S. B., & Dubs, A. (1982). Predictive coding: a fresh view of inhibition in the retina. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *216*(1205), 427–459.
Tootell, R. B., Hamilton, S. L., & Switkes, E. (1988). Functional anatomy of macaque striate cortex. IV. Contrast and magno-parvo streams. *Journal of Neuroscience*, *8*(5), 1594–1609.

Torralba, A., & Oliva, A. (2003). Statistics of natural image categories. *Network: Computation in Neural Systems*, *14*(3), 391–412.

Tusa, R. J., Palmer, L. A., & Rosenquist, A. C. (1978). The retinotopic organization of area 17 (striate cortex) in the cat. *Journal of Comparative Neurology*, *177*(2), 213–235.

Ullman, S. (1996). *High-level vision: Object recognition and visual cognition* (Vol. 2). MIT press Cambridge, MA.

Van Essen, D. C., Anderson, C. H., & Felleman, D. J. (1992). Information processing in the primate visual system: an integrated systems perspective. *Science*, *255*(5043), 419–423.

Van Essen, D. C., & Maunsell, J. H. R. (1983). Hierarchical organization and functional streams in the visual cortex. *Trends in Neurosciences*, *6*, 370–375.

Vassilev, A., & Mitov, D. (1976). Perception time and spatial frequency. *Vision Research*, *16*(1), 89–92.

Wandell, B. A. (1995). Foundations of Vision (Sinauer, Sunderland, MA). *WandellFoundations of Vision1995*.

Weinsberg, U., Li, Q., Taft, N., Balachandran, A., Sekar, V., Iannaccone, G., & Seshan, S. (2012). CARE: content aware redundancy elimination for challenged networks. In *Proceedings of the 11th ACM workshop on hot topics in networks* (pp. 127–132). ACM.

Welch, T. A. (1984). A technique for high-performance data compression. *Computer*, (6), 8–19.

Zeki, S. M. (1978). Uniformity and diversity of structure and function in rhesus monkey prestriate visual cortex. *The Journal of Physiology*, *277*(1), 273–290.

Ziv, J., & Lempel, A. (1978). Compression of individual sequences via variable-rate coding. *IEEE Transactions on Information Theory*, *24*(5), 530–536.

Chapter 2

Spatial frequency tuning to single and composite sinewave gratings

Kirsten Petras[1,2,*], Alexandra Vossen[3,*], Valerie Goffaux[1,2,3,*], Sarang S. Dalal[4,*]

1 Research Institute for Psychological Science, Université Catholique de Louvain, Louvain-la-Neuve, Belgium; 2 Department of Cognitive Neuroscience, Maastricht University, Maastricht, the Netherlands; 3 Institute of Neuroscience, Université Catholique de Louvain, Louvain-la-Neuve, Belgium; 4 Center of Functionally Integrative Neuroscience, Aarhus University, Aarhus, Denmark
* Authors contributed equally

Abstract

One of the fundamental properties of visual images is their spatial frequency content. At the earliest stages of visual processing (from the retina to striate cortex), neurons show highly selective responses for narrow spatial frequency bands, and spatial frequency tuning functions have been described using single frequency gratings. Higher order "object-selective" visual areas on the other hand tend to respond preferentially to more complex visual patterns that share the statistical properties of the area's preferred object category. It is therefore plausible that spatial frequency tuning properties in those areas are better addressed using stimuli that conform to the area's preferences. Here, we use MEG to compare sensor and source-space responses to both single frequency sinewave gratings and narrowband noise patches, that share the characteristic $1/f^{\alpha}$ amplitude spectrum of natural images, at different stages of the ventral visual pathway. We found spatial frequency-specific responses in visually responsive areas outside the occipital lobe, indicating that higher-level visual areas are still sensitive to primary stimulus properties. Contrary to our expectation however, we found no difference between responses to single sinewave and narrowband gratings. This suggests that spatial frequency tuning is preserved even at higher levels of the visual hierarchy but might be independent of stimulus complexity.

Introduction

The contrast sensitivity function relates the spatial frequency of a grating to the minimum contrast at which it can be detected by a human observer. It can be defined using a large variety of methods, ranging from animal invasive recordings, human psychophysics (Georgeson & Sullivan, 1975), electroretinogram (ERG) (Korth, Rix, & Sembritzki, 1985), electroencephalography (EEG) (Norcia, Tyler, Hamer, & Wesemann, 1989) recordings, and metabolic measures such as fMRI (Tootell et al., 1998) and fNIRS (Plichta, Heinzel, Ehlis, Pauli, & Fallgatter, 2007). Single sinewave Gabor patches, produced by convolving a sinusoid with a Gaussian, have customarily been used to quantify CSF. CSF is thought to primarily reflect the function of primary visual cortex (V1) neurons, on account of their selectivity to narrow ranges of spatial frequency and orientation (B. W. Andrews & Pollen, 1979; De Valois, De Valois, & Yund, 1979). In contrast, downstream regions along the ventral visual pathway or visually responsive frontal areas with their larger and more complex receptive fields have traditionally been thought to selectively respond to a range of object categories and visual scenes (Aguirre, Zarahn, & D'esposito, 1998; Ishai, Ungerleider, Martin, Schouten, & Haxby, 1999; Murata et al., 1997; Peelen & Downing, 2017; Schwarzlose, Swisher, Dang, & Kanwisher, 2008) and are often experimentally targeted with naturalistic images. However, multiple areas within the ventral visual pathway have been shown to also prefer stimuli with statistical properties similar to their preferred object or scene category. The fusiform face (FFA) area for example not only shows stronger responses to faces than other visual categories such as e.g. cars but also portrays a similar preference for phase-scrambled images of the same categories (Rossion, Hanseeuw, & Dricot, 2012). Andrews et al. (2010) demonstrated a response bias of face- and scene-selective areas towards images with similar spatial frequency and orientation spectra as their respective preferred category. Other studies confirmed a sensitivity to the spatial frequency and orientation content of natural scenes and faces in their respective selective regions (Goffaux, Duecker, Hausfeld, Schiltz, & Goebel, 2016; Goffaux et al., 2010; Kauffmann, Ramanoël, Guyader, Chauvin, & Peyrin, 2015; Nasr & Tootell, 2012; Watson, Hartley, & Andrews, 2014). Those findings indicate that the spatial frequency spectrum and orientation profile of a stimulus are sufficient to elicit differential responses from category-selective high-level areas. To understand how those higher-level visual regions retain sensitivity to basic features of visual stimuli, it is therefore important to more progressively map the stimulus space between single sines and fully naturalistic object stimuli.

One of the most fundamental properties of natural images is their characteristic amplitude spectrum, in which power decreases linearly with spatial frequency on a log-log scale (Van der Schaaf & van Hateren, 1996), producing a characteristic $1/f^{\alpha}$ power spectrum with α between 1.2 and 2.3 for most image categories (Burton &

Moorhead, 1987; Field, 1987; Ruderman & Bialek, 1994; Tolhurst, Tadmor, & Chao, 1992; Van der Schaaf & van Hateren, 1996). Hence, as a first step towards more naturalistic stimuli, while retaining a high level of control over stimulus properties, we generated grating stimuli with a naturalistic $1/f^{\alpha}$ amplitude spectrum. With this manipulation, we produce images that lack any meaningful shape, thus allowing us to attribute neural responses purely to the fundamental image properties of spatial frequency and amplitude spectrum. In contrast to single sinewave gratings, composite sinewave gratings, i.e. gratings produced by the superposition of more than one sinewave, can be adjusted to produce stimuli that are centered on a single spatial frequency and orientation while following a $1/f^{\alpha}$ power drop off. Given that $1/f^{\alpha}$ composite sinewave gratings show a higher statistical similarity to natural images than single sinewave gratings, it could be expected that they are superior in their ability to activate higher-level visual regions and therefore provide sufficiently strong activation for the estimation of contrast sensitivity functions further beyond V1 than single sinewave gratings.

In order to address this question, we employed a steady state visual evoked potential (SSVEP) paradigm during which participants viewed sequences of flickering stimuli that contained either single sinewave Gabor patches, or $1/f^2$ narrowband composite sinewave gratings with the same center spatial frequencies and orientations. During stimulus presentation, we measured MEG as well as ERG to be able to track responses of neuronal populations along the ventral visual pathway.

We hypothesize that response amplitudes for single sinewave gratings decay faster with distance from V1 towards the ventral visual pathway compared to those evoked by $1/f^2$ composite sinewave gratings. We further posit, that spatial frequency, as a fundamental stimulus property, is represented in higher-level visual areas and can be retrieved via multivariate response classification.

Methods

Eleven healthy volunteers (mean age 30.55y, 4 females) were recruited for the experiment. All participants provided written informed consent and procedures were approved by the local ethical committee (Approval Nr: 1-10-72-21-16). Two participants were excluded from further analysis. One due to technical errors and another one due to excessive blinking and poor data quality. If participants required optical correction, lenses matching their prescription were glued directly to the MEG helmet.

Stimuli

A total of 40 basis images spanning 2417x2417 pixels were created. Twenty of them, further referred to as single-sinewave gratings, contained a single 2D

sinewave each, and 20 of them, further referred to as narrowband gratings, contained a combination of sinewaves, centered on the same spatial frequencies and orientations as the single-sinewave gratings. Sinewave gratings were created at initially six spatial frequencies, logarithmically spaced between 0.6 and 10 cycles per degree of visual angle (cpd), and four orientations (0°, 45°, 90° and 135°) with random phase. To reduce recording time, we later chose to drop the highest SF.

The creation of the narrowband gratings largely followed the procedure described in Haun & Essock (2010). In brief, for each of the center spatial frequency by orientation combinations, composite sinewave gratings were produced as follows: A set of equal contrast, random phase 2D sinewaves were created at 6 spatial frequencies (0.2 octaves apart from each other, centered on the spatial frequency of the corresponding single-sinewave gratings) and at 15 orientations (center orientation ± 21° in 7 steps of 3°) per spatial frequency. The narrowband gratings were then assembled by adding the resulting 90 individual 2D sinewaves per image. The 0.2 octave spatial frequency steps ensured that the lowest and the highest spatial frequency contained in each narrowband image were 1 octave apart. Because in images whose amplitude spectrum shows a $1/f^2$ characteristic each octave of spatial frequencies contributes equally to root-mean-square (RMS) contrast, our one octave narrowband stimuli are similarly matched in RMS contrast across center SF conditions. While a similar result could in theory be achieved by multiplying white noise images with Fourier space filters that restrict the 2D amplitude spectrum to the desired spatial frequency and orientation ranges, such as the example shown in figure 1B, we chose for the addition of equal contrast sinewave gratings in native image space to avoid orientation inaccuracies at low spatial frequencies (Hansen & Essock, 2004).

Stimulus generation codes will be made freely available on github (https://github.com/kpetras/SF_flicker). The final set of stimuli consisted of 14 gratings per trial, spanning 1209x1209 pixels (20° of visual angle) that were each sampled with random starting points from the basis image of the corresponding condition (see figure 1C for illustration). This random sampling ensured that grating phase varied randomly across images, while at the same time being computationally more efficient than calculating each random phase image individually. Gratings were then convolved with a cosine-windowed disk-shaped aperture such that grating contrast decreased with eccentricity starting at a radius of 9 degrees from fixation until it reached 0 at 10 degrees of eccentricity (see figure 1D). The contrast of the single sinewave gratings was reduced to match that of the narrowband gratings (~ 0.11) and all images were normalized to a mean of 0.5. Thus, while the histograms differed considerably between single sinewave and narrowband gratings, RMS contrast and mean luminance were comparable.

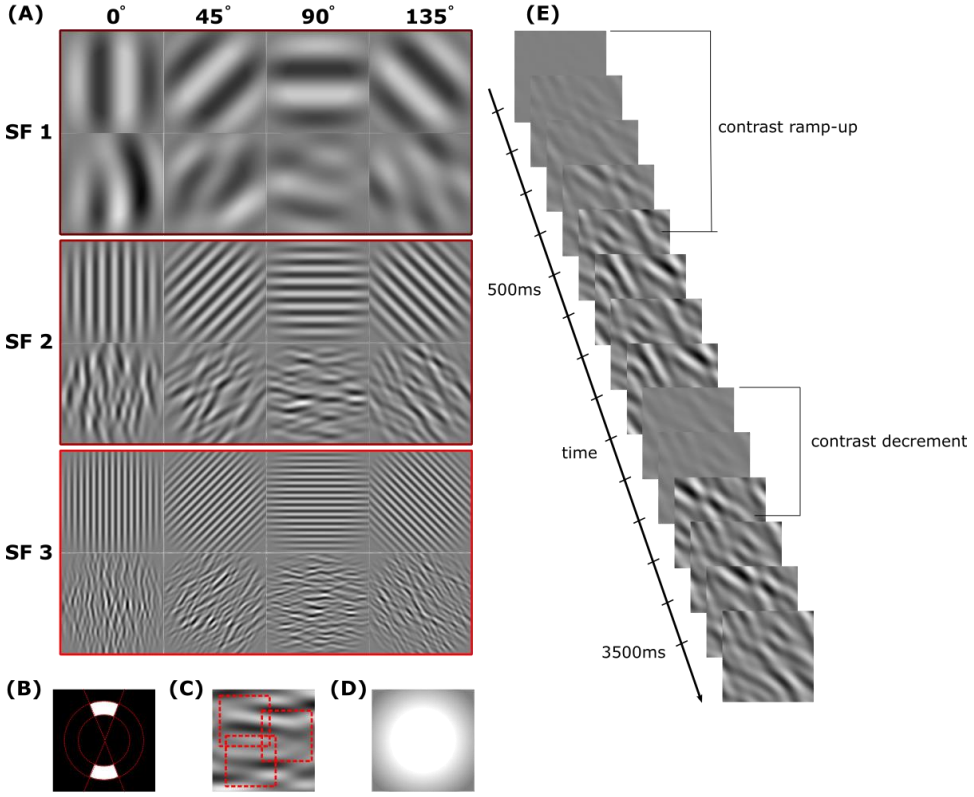


Figure1: (A) Example stimuli for 3 SF steps. Narrowband gratings are centered on the same spatial frequency and orientation as the corresponding sinewave gratings. (B) Corresponding Fourier filter. The passband of spatial frequencies forms an annulus that intersects with a wedge formed by the orientation spread. (C) Random cutouts of the larger basis image were chosen to obtain random phase gratings *(D) Gratings were tapered off with a circular mask to avoid oriented edges at the outlines of the stimulus. (E) Example sequence (task trial). Grating contrast increases to its maximum level during the initial ramp-up period. Participants were asked to report a contrast decrement (see middle two gratings) via a button press.*

All images were generated in Matlab and presented via the Psychtoolbox (Brainard, 1997) on a ProPixx projector (VPixx Technologies, Canada) in M16 mode, enabling 10-12 bits of grayscale resolution at a monitor refresh rate of 120 Hz and a mean luminance of 350 cd/m2.

Procedure

All recordings were acquired at the Center of Functionally Integrative Neuroscience, Aarhus University, Denmark.
The electroretinogram (ERG) was recorded from both eyes via corneal silver/nylon electrodes (DTL Plus Electrode, Diagnosys, USA) referenced to the ipsilateral outer canthi. We found no periodic response in the ERG, likely due to the low contrast of our stimulation. ERG data was therefore not analyzed further. Simultaneously, MEG (Elekta Neuromag TRIUX) was recorded from 102 magnetometers and 204 orthogonal planar gradiometers at a sampling rate of 2000Hz in a magnetically shielded room. If visual acuity needed to be corrected, a lens matching the subject's prescription was attached directly to the MEG helmet in front of the affected eye.
Participants were instructed to fixate a central dark gray dot and press a button whenever they detected a contrast decrement in the stimulus. Each participant completed 400 trials (5 spatial frequencies x 4 orientations x 2 grating types x 10 repetitions). Each trial consisted of a 3.5 second period during which 14 gratings were presented at a rate of 8 Hz in a 50% duty cycle, resulting in an effective grating onset frequency of 4 Hz. Gratings within a trial were of a given bandwidth (i.e. single sinewave or narrowband), (center) spatial frequency and (center) orientation but with a random phase shift between successive gratings. During the first 500 ms grating contrast gradually ramped up to the maximum RMS contrast of 0.11 to avoid strong onset responses in the ERG/MEG that would contaminate the subsequent periodic response (SSVEP; see data analysis). Condition order was pseudo-randomized such that no more than three trials of the same spatial frequency, orientation or grating type could appear in direct succession to avoid potential adaptation effects. The inter-trial interval was jittered between 1.5 and 2.5 seconds. 40 task trials, in which participants were asked to detect a brief contrast decrement, were equally distributed across conditions. In those trials, after the initial contrast ramp-up, contrast was decreased by 20% for two consecutive gratings starting at least three gratings before the end of the trial. After half of every block of 40 trials, participants had a five second break during which a countdown announced the time to the start of the next trial; at the end of each block they could take a self-timed break to rest their eyes.
To facilitate subject specific head models for source reconstruction, all participants underwent an MRI scanning session (3 T Siemens Skyra) consisting of a high resolution, full brain T1 weighted scan (1 mm isotropic, reconstructed matrix size 256 × 256) as well as a T2 weighted scan. Total scan duration was about 25 min. MRIs were converted to nifti format and aligned to AC/PC space using SPM 12 (Penny, Friston, Ashburner, Kiebel, & Nichols, 2011).

MEG preprocessing

MEG Data were band pass filtered between 0.1 and 100Hz using a Hamming windowed zero-phase FIR filter and down-sampled to 1000Hz in MNE python version 0.16.1 (Gramfort et al., 2014). Subsequent analysis was performed in Matlab 2019a using functions from the fieldtrip toolbox (Oostenveld, Fries, Maris, & Schoffelen, 2011) as well as custom scripts. All scripts are made freely available on GitHub (https://github.com/kpetras/SF_flicker). Power line noise (50Hz + first harmonic) was removed via spectrum interpolation (Leske & Dalal, 2019). Data were segmented into 3.5 second epochs starting with flicker onset and ending with flicker offset. Trials during which the vertical EOG exceeded a trial-wise Z-score of 3 and absolute amplitude exceeded 50µV were marked as artifactual and excluded from further analysis.

Sensor space frequency domain analysis

Presenting stimuli at a fixed temporal frequency gives rise to the steady state visual evoked potential, a neural response with a periodicity matching the stimulation frequency (see Norcia, Appelbaum, Ales, Cottereau, & Rossion, 2015 for review). To take full advantage of this periodicity in the neural response, we analyzed our data in the frequency domain. We calculated discrete Fourier transforms of the average time series waveform per participant and spatial frequency, pooled over all four orientations. The 3.5 s segments resulted in a frequency resolution of 0.2857 Hz. We analyzed responses for all posterior MEG channels. In accordance with previous literature, we first established the number of harmonics of the 4 Hz stimulation frequency that significantly exceeded noise levels in the grand average amplitude spectrum. To this end, grand average spectra of all trials and participants were converted into z-scores by subtracting, for each frequency bin, the mean amplitude of the 6 surrounding bins (three to either side) and dividing the result by the standard deviation of the surrounding bins (see Quek, Liu-Shuang, Goffaux, & Rossion, 2018 and figure 1F). Harmonic frequency bin amplitudes were significantly above noise amplitudes for up to 56 Hz at a threshold of $z > 2.57$ ($p \leq 0.005$, one-tailed hypothesis i.e. signal > noise). To avoid any influences of residual power line noise, we limited further analysis to frequencies up to 40 Hz. To obtain a single measure of the strength of the periodic response, the Z-scores at the stimulation frequency and its harmonics were summed per condition. Differences between conditions were statistically assessed with repeated measures ANOVAs of this periodic response measure, averaged over all posterior channels (N=72), with the factors grating type (levels: single sinewave, narrowband) and spatial frequency (SF; levels 1-5).

For visualization in native units (figure 2), the amplitudes at the 4 Hz stimulation frequency as well as its harmonics up to 40 Hz were baseline-corrected by

subtracting the mean of the six surrounding bins and subsequently summed for each participant and condition.

Source space analysis

Individual headmodels were constructed based on each participant's MRI. The MEG data were aligned to MRI space via a combination of fiducial point matching and headshape matching using custom code as well as fieldtrip (Oostenveld et al., 2011) functions. Fiducial positions and approximately 250 headshape points were acquired with the Polhemus Fastrak® digitizer. The participants' head location in the MEG helmet was recorded via four head position indicator coils (two placed on the forehead and one on each mastoid).
Individual MRIs were segmented into gray matter, white matter, CSF, skull and scalp and the mid gray matter cortical sheet was extracted in freesurfer software (freely available here: http://surfer.nmr.mgh.harvard.edu). To obtain a computationally efficient, but still highly accurate surface mesh, the freesurfer boundary meshes were reduced to ~15,500 vertices using the HCP workbench pipelines (Marcus et al., 2011). Because the HCP pipeline produces cortical meshes that are surface-registered to a common template, dipole locations with the same index number correspond to equivalent coordinates across subjects and can therefore be directly compared without the need for further interpolation. For each participant, a single shell headmodel was created based on the reconstruction of the pial surface provided by freesurfer. Leadfields for ~15,500 equivalent current dipole positions based on the individual participants' cortical sheet were calculated using the OpenMEEG implementation in fieldtrip (Gramfort, Papadopoulo, Olivi, & Clerc, 2010; Kybic et al., 2005; Oostenveld et al., 2011). The sources of periodic neural responses to visual stimulation were reconstructed using minimum norm estimates (MNE) in fieldtrip with noise covariance scaling. Noise covariance scaling accounts for the fact that the noise covariance is calculated on a single trial level prior to averaging, thus not reducing noise over trials, while the data used to calculate the source analysis is averaged first, thus much of the noise averages out. To obtain the noise covariance matrix, we exploited the fact that the periodic stimulation allows us to define the response at the stimulation frequency and its harmonics as the signal of interest. Accordingly, the activity at any other frequency qualifies as noise. We therefore constructed noise covariance matrices from the data as follows: For each trial, we used spectrum interpolation (Leske & Dalal, 2019) to replace the power at the frequencies of interest (i.e. the signal) with "noise", i.e. the mean of the 3 neighboring frequency bins to either side skipping the directly adjacent ones. This method preserves the original phase information, but effectively removes the response of interest. The data were then transformed back into the time domain using inverse Fourier transform. The final noise covariance

matrix is the scaled average (using the scaling factor lambda = 3) of the single trial noise covariances estimated on the spectrum-interpolated time series.

From the reconstructed source time series, we calculated the discrete Fourier transform for each subject, dipole, and condition separately. As in sensor space, the resulting source power spectra at the stimulation frequency and its first nine harmonics up to and including 40 Hz were z-transformed to the activity at their respective neighboring frequencies, and then summed to obtain an estimate of the overall periodic response.

To address our hypothesis that narrowband $1/f^2$ noise will evoke responses further beyond primary visual cortex than single sinewave gratings do, we further parcellated the source space into regions of interest (ROIs) based on gyral and sulcal structure of the anatomical MRI using the Desikan-Killiany atlas (Desikan et al., 2006) implemented in freesurfer. We selected six ROIs for further analysis: pericalcarine, lateral occipital, fusiform, parahippocampal, lateral orbitofrontal, and medial orbitofrontal. ROIs were selected to cover the occipital and the ventral visual pathway as well as orbitofrontal cortex since those regions have been implicated in visual object and scene processing (e.g. Grill-Spector, 2003; Pietrini et al., 2004). The selected ROIs are shown in figure 2.

To reduce the influence of less "responsive" sources, we selected for each participant separately the 50% most active sources (i.e. median cutoff) per ROI based on the respective grand average periodic response across all stimulus conditions. The median of these selected sources was entered for statistical analysis into a repeated measures ANOVA with factors ROI (pericalcarine, lateral occipital, fusiform, parahippocampal, lateral orbitofrontal, medial orbitofrontal), spatial frequency (SF; five levels), and grating type (single sinewave, narrowband).

All ANOVA test statistics were evaluated at alpha = 0.05. In comparisons of variables with more than two levels, degrees of freedom were adjusted using the Greenhouse-Geisser correction to account for violations of sphericity. In Post-hoc t-tests for significant main effects and interactions, the p-values were adjusted using Bonferroni correction to reduce the chance of type I errors due to multiple comparisons.

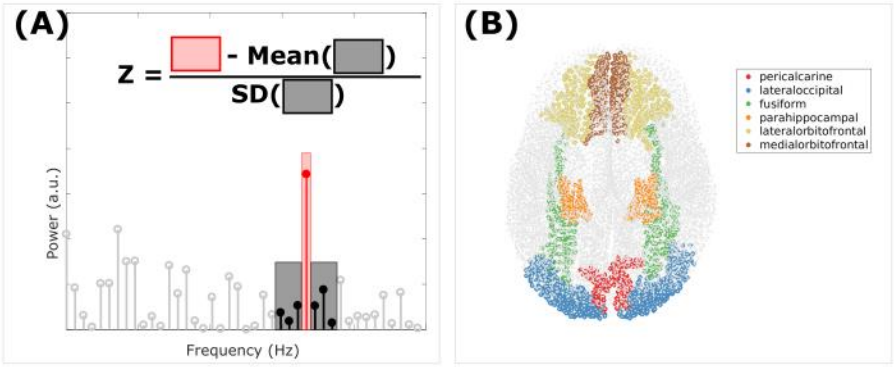


Figure 2: A) For each frequency bin, Z-scores were calculated by subtracting the mean of the neighboring frequency bins (3 to either side) and dividing the result by the standard deviation of the neighboring frequency bins B) Regions of interests (ROIs) for source analysis. Each circle represents a vertex in the individual source model. Colors correspond to the respective parcellations in the Desikan-Killiany atlas (dorsal view). ROIs are associated with different stages of visual scene processing along the ventral visual pathway up to visually responsive orbitofrontal areas.

Spatial frequency classification

To investigate whether the spatial pattern of source activation varies with stimulus spatial frequency and differs between grating types, we used multiclass linear discriminant analysis (mcLDA), implemented in the MVPA-light package (https://github.com/treder/MVPA-Light), to classify individual trials according to which spatial frequency was presented. LDA projects data components onto a lower dimensional space, while maximizing class separation in the projection onto component axes (Rao, 1948).

To focus on multivariate differences, we removed the univariate difference between conditions by subtracting, for each participant, condition and ROI, the mean response over trials and vertices prior to classification. The first twenty principle components (PCs) in descending order of their contribution to the data variance were selected as features for the mcLDA.

To establish whether PCs were separable between SF conditions, we first ran a *within grating type* classification per participant and ROI, where the classifier was both trained and cross-validated on trials belonging to the same grating type condition. We used five-fold cross validation (data is randomly split into folds with part of the data held out for cross-validation) and 10 repetitions (new random splits). Second, we ran an *across grating type* classification to test whether the features used to classify one stimulus type are general enough to yield comparable classification accuracy for the other stimulus type. This classifier was trained using

all available data of one grating type and then evaluated on all data from the other grating type condition (i.e. trained on single sinewave, tested on narrowband gratings (TrainSingleTestNarrow) and vice versa (TrainNarrowTestSingle)). Classifier performance was defined as the average accuracy over classification folds and repetitions for within grating type classification, and the average over TrainSingleTestNarrow and TrainNarrowTestSingle for across grating type classification. Thus, we compared the performance of three different classifiers: within single sinewave, within narrowband, and generalized across grating type.
To evaluate the statistical significance of the classification result, we calculated (for each subject, ROI, and classifier) a permutation distribution on trials with randomly re-assigned category labels (5000 permutations) to derive an empirical chance distribution. We considered a classifier's performance as significantly above chance if it was outside the 97.5th percentile of its respective empirical chance distribution. The results for each classifier are corrected for multiple comparisons using the Bonferroni method, by dividing the one-sided (classification accuracy > chance) alpha level (0.025) by the number of ROIs. We then compared classification performance across ROIs and grating types with a repeated measures ANOVA with factors ROI (pericalcarine, lateral occipital, fusiform, parahippocampal, lateral orbitofrontal, medial orbitofrontal), and classification type (within single sinewave, within narrowband, generalized across grating type).
All ANOVA test statistics were evaluated at alpha = 0.05. Degrees of freedom were adjusted using the Greenhouse-Geisser correction. In post-hoc t-tests for significant main effects and interactions, the *p*-values were adjusted using Bonferroni correction.

Results

Sensor space steady state responses show spatial frequency tuning

Figure 3 shows the grand average amplitude spectrum over all (gradiometer) sensors, averaged across spatial frequencies. All amplitudes are baseline-corrected to the neighboring frequency bins as described above. Neural responses at the stimulus presentation frequency of 4 Hz exceeded noise levels for all participants and most spatial frequency conditions.

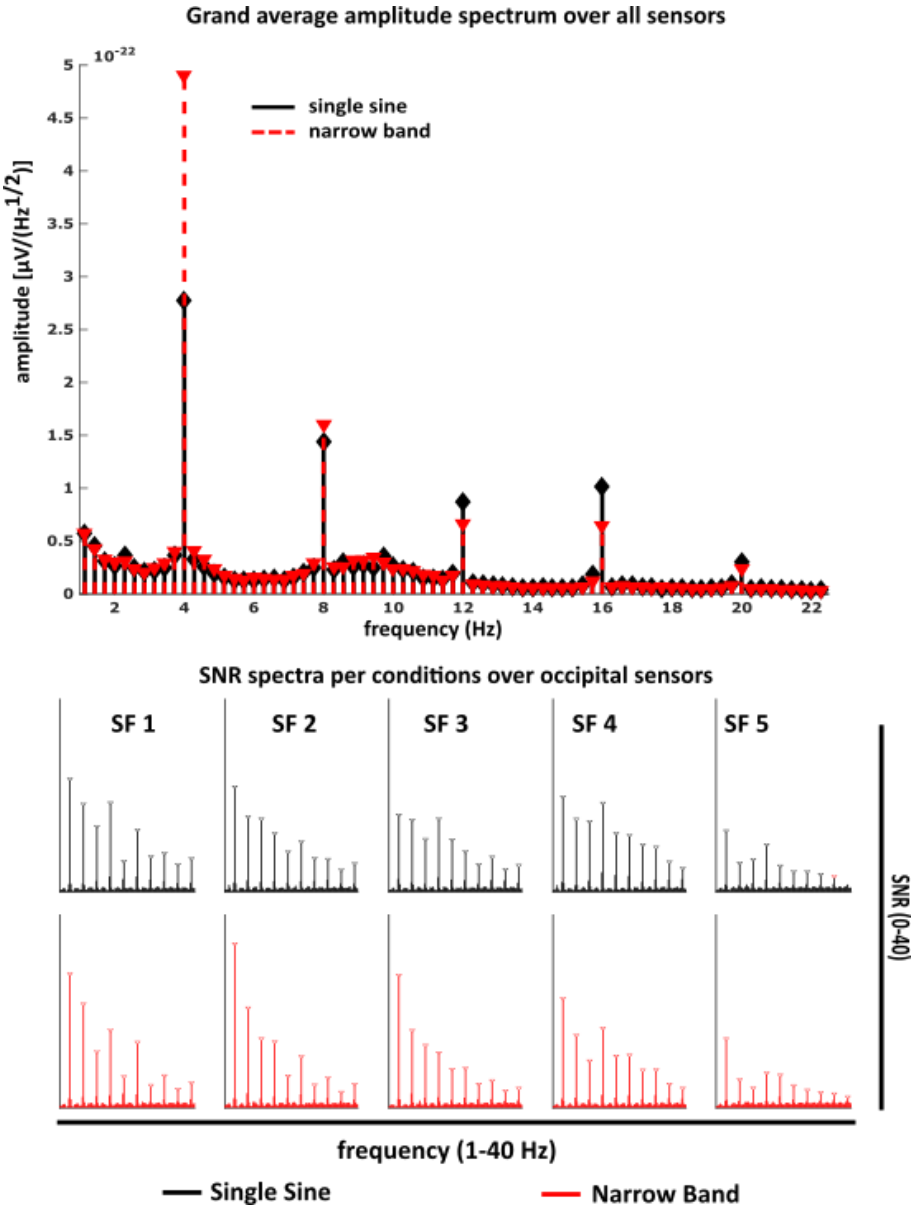


Figure 3: Baseline corrected grand average amplitude spectrum (top) and individual condition SNR spectra *(bottom).* The summed flicker response (4Hz and harmonics) exceeded noise levels for all conditions. The maximum SNR at a single frequency (4Hz) was 37 (SF2, narrow band).

Z-transformed, summed harmonic responses to narrowband gratings over posterior channels were generally more pronounced than those to single sinewave gratings [repeated measures ANOVA, $F(1, 8) = 7.15$; $p < .05$]. Stimulus center spatial frequency significantly modulated responses [repeated measures ANOVA, $F(1.60, 12.77) = 20.57$, $p < .001$, $\varepsilon = .40$]. Both responses to single sinewave and to narrowband gratings showed spatial frequency tuning consistent with previous studies, with higher sensitivity to intermediate spatial frequencies (see figure 4). The maximum response, on average over all participants, was found for SF step four (SF = 4.95 cpd) for both stimulus types (single sinewave and narrowband), however, responses showed considerable variability across participants (see error bars figure 4A). Based on visual inspection of the harmonic responses to the different stimulus conditions (see figure 3, bottom panel), we post-hoc tested whether the pattern of harmonics differed between single sinewave and narrowband gratings. Repeated measures ANOVA with factors ROI (pericalcarine, lateral occipital, fusiform, parahippocampal, lateral orbitofrontal, medial orbitofrontal), grating type (single sinewave, narrowband) and harmonic frequency (4-40Hz, 10 steps) revealed no significant interaction between grating type and harmonic frequency (p>0.05). We are therefore confident that the measure we used in our original analysis, i.e. summed harmonic responses, that are invariant to the pattern of the individual harmonics, is sufficient to describe our data.

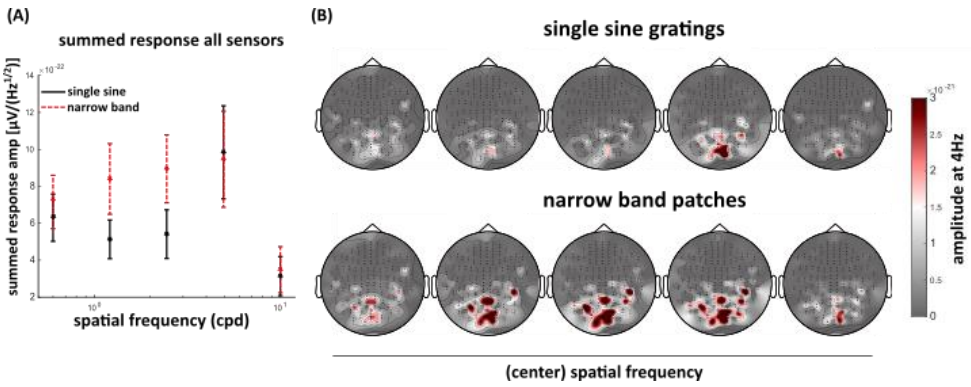


Figure 4: (A) Grand average summed response amplitudes averaged over all sensors. Error bars indicate standard error of the mean (B) topographical representation of 4 Hz response amplitude. Narrowband patches consistently elicit a stronger response than single sinewave gratings.

Source reconstructed responses peak in occipital cortex

We hypothesized that, despite the absence of a meaningful shape, narrowband $1/f^2$ noise activates regions along the ventral pathway beyond primary visual cortex more strongly than single sinewave gratings. Figure 5 shows the source reconstructed activation, averaged over all participants, projected onto the inflated reconstruction of the cortical surface of an example participant.

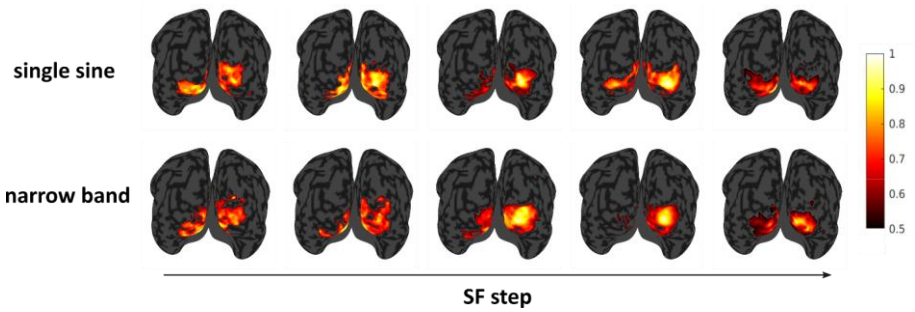


Figure 5: caudal view of the inflated surface reconstruction of an example participant with the source reconstructed activation (averaged over all participants) projected. Activation is thresholded at Z > 2.5 standard deviations from the mean and plotted in arbitrary units.

Figure 6 shows the median periodic response for single sinewave and narrowband gratings at each spatial frequency and in each ROI. The strongest response across all ROIs was elicited by gratings with a spatial frequency of 4.95 cpd, and the lowest response by stimuli patterned at 10 cpd. Especially for the lower spatial frequencies of up to 5 cpd, narrowband composite sinewave gratings elicited, on average, slightly stronger periodic responses in line with our hypothesis. However, this pattern repeats across all ROIs. Overall, responses in early visual cortex were larger than in higher-level visual areas. In line with these observations, the ANOVA shows significant main effects of spatial frequency ($F(2.14,17.11) = 11.62$, $p < .0005$, $\varepsilon = .53$), and ROI ($F(2.35,18.88) = 9.5$, $p = .0009$, $\varepsilon = .47$). There was no significant interaction between SF and ROI ($p > 0.05$).

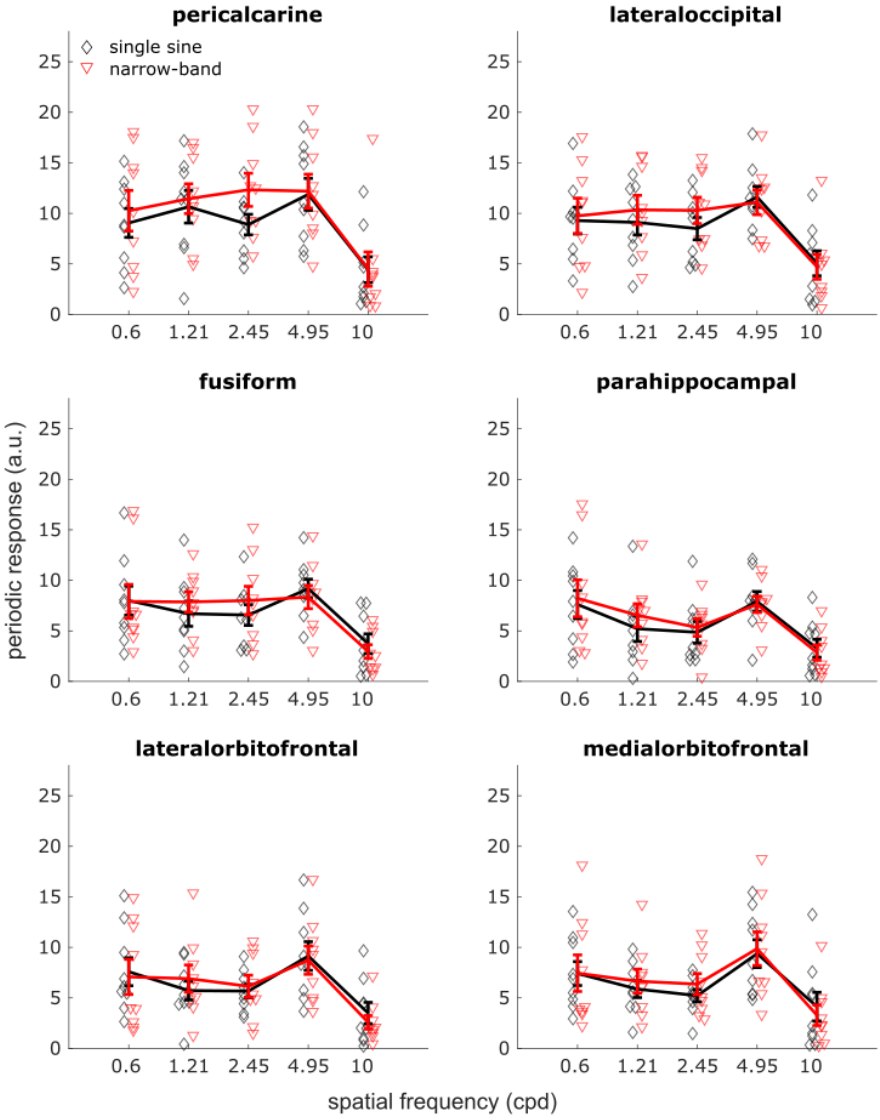


Figure 6: Periodic response to single sinewave vs. narrowband gratings of different spatial frequencies in different brain regions of interest along the visual hierarchy. Periodic response is estimated by the z-score normalized power at the stimulus presentation frequency and its first nine harmonics (see main text for details). Spatial frequency is reported in cycles per degree of visual angle (cpd). Individual data (N = 9) are summarized as mean and standard error. While stronger responses were elicited in early visual areas, the same pattern of spatial frequency tuning can be observed in all ROIs and is statistically equivalent between single sinewave (black diamonds and lines) and narrowband (red triangles and lines) stimuli. (a.u. = arbitrary units).

Post-hoc tests for spatial frequency purport that stimuli at the highest spatial frequency level elicited a smaller periodic response than stimuli at levels 2-4 (SF2 vs. SF5: $t(8) = 3.84$, $p = .0050$; SF3 vs. SF5: $t(8) = 3.70$, $p = .0061$, SF4 vs. SF5: $t(8) = 3.94$, $p = .00001$, alpha level corrected for multiple comparisons = .0086), but was similar to the lowest spatial frequency (all other comparisons $p > .009$).
Post-hoc tests for ROI show that the periodic response in pericalcarine cortex was stronger than in the parahippocampal gyrus ($t(8) = 3.84$, $p = .0050$) and both lateral ($t(8) = 4.29$, $p = .0027$) and medial orbitofrontal cortex ($t(8) = 3.91$, $p = 0.0045$; corrected alpha = .0065 for all Post-hoc comparisons for ROI). In addition, the periodic response was stronger in lateral occipital cortex than in both parahippocampal gyrus ($t(8) = 4.06$, $p = .0037$) and lateral orbitofrontal cortex ($t(8) = 3.73$, $p = .0058$). Finally, the response in the fusiform gyrus was statistically higher than in the parahippocampal gyrus ($t(8) = 3.69$, $p = .0061$). So overall upstream visual areas show a stronger periodic response than more down-stream areas. While none of the other ROI comparisons meet the significance threshold, the general trend is towards decreasing responses with increasing hierarchical level (see figure 7B). But given that Euclidian distance between pericalcarine, where the strongest response was observed, and the other ROIs differed, field spread must be considered as an alternative explanation (see discussion).
Importantly for our hypothesis, however, neither the interaction between grating type and ROI ($F(2.09, 16.75) = .74$, $p = .497$, $\varepsilon = .42$), nor the three-way interaction of grating type, ROI, and spatial frequency ($F(3.82, 30.53) = 0.39$, $p = .808$, $\varepsilon = .19$) were significant. In other words, we found no statistical evidence that the more naturalistic narrowband $1/f^2$ stimuli elicit a relatively stronger response, compared to simple sinewave gratings, in higher visual areas. None of the other tests showed statistical differences, although the *p*-values for the interaction of ROI and spatial frequency ($F(4.82, 38.55) = 2.35$, $p = .061$, $\varepsilon = .24$, and the main effect of grating type ($F(1, 8) = 3.95$, $p = .0821$), were close to our chosen alpha level, the latter reflecting the on average somewhat greater response for narrowband stimuli also seen in the sensor space data. However, as the strength of the statistical evidence is weak and these comparisons are immaterial to our hypothesis, they will not be discussed in more detail.

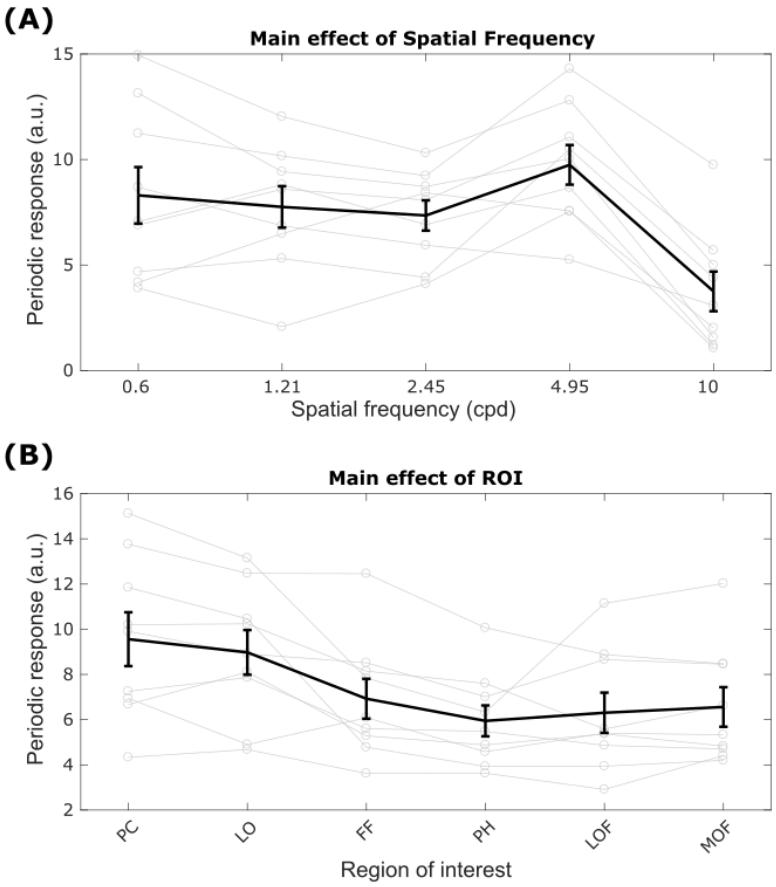


Figure 7: Main effects of (A) spatial frequency and (B) ROI. Grey lines with open circles represent individual participants, black lines represent mean and standard error. Periodic response is estimated by the z-score normalized power at the stimulus presentation frequency and its first nine harmonics (see main text for details). Spatial frequency is reported in cycles per degree of visual angle (cpd). (A) Gratings at around 5 cpd elicited stronger periodic responses compared to lower and higher spatial frequencies. (B) Early visual areas showed a stronger response to flickering gratings compared to more downstream areas. Abbrev.: PC = pericalcarine, LO = lateral occipital, FF = fusiform, PH = parahippocampal, LOF= lateral orbitofrontal, MOF = medial orbitofrontal.

Spatial patterns of source activation differ between grating types

The univariate approach did not reveal any changes in the relative preference for more naturalistic narrowband gratings along the ventral pathway. However, the spatial pattern of responding sources *within* each ROI could potentially

differentiate between grating types, if different neuronal populations within an area are sensitive to different stimulus aspects. To investigate whether spatial frequency representation depends on grating type, we used LDA to classify trials belonging to different spatial frequency categories within and across grating type for each ROI.

All classifiers were able to predict grating spatial frequency significantly better than empirical chance in all ROIs (see figure 8). Both for single sinewave gratings and for narrowband gratings, mean classification accuracies were above the 97.5^{th} percentile of the permutation distribution. Classifier generalization across grating type (training on single sinewave and evaluating on narrowband gratings and vice versa) also resulted in accuracies that were significantly above empirical chance in all ROIs (mean values outside 97.5^{th} percentile of permutation distribution).

To evaluate potential differences between ROIs and classification types (within grating type or across grating type), we calculated a repeated measures ANOVA with factors ROI (pericalcarine, lateral occipital, fusiform, parahippocampal, lateral orbitofrontal, medial orbitofrontal) and classification type with two levels: within grating type (classifier accuracy averaged across grating types), and generalized across grating types. The ANOVA indicated a main effect of ROI ($F(2.03, 16.25) = 7.39$, $p = .005$, $\varepsilon = .41$). Post-hoc contrasts at a corrected alpha of .006 showed differences between the lateral occipital ROI and the pericalcarine ($t(8) = 4.02$, $p = .004$), the parahippocampal ($t(8) = 3.86$, $p = .005$) and the medial orbitofrontal ($t(8) = 3.99$, $p = .004$) ROIs with higher accuracies for lateral occipital ROI.

We further found a significant main effect of classification type ($F(1, 8) = 10.11$, $p = .013$). The contrast Within – Across indicates that classification performance was consistently higher when classifiers were trained and tested on data from the same grating type condition, rather than trained and evaluated on different grating type conditions ($t(8) = 3.18$).

We found no significant interaction between classification type and ROI ($p > .05$).

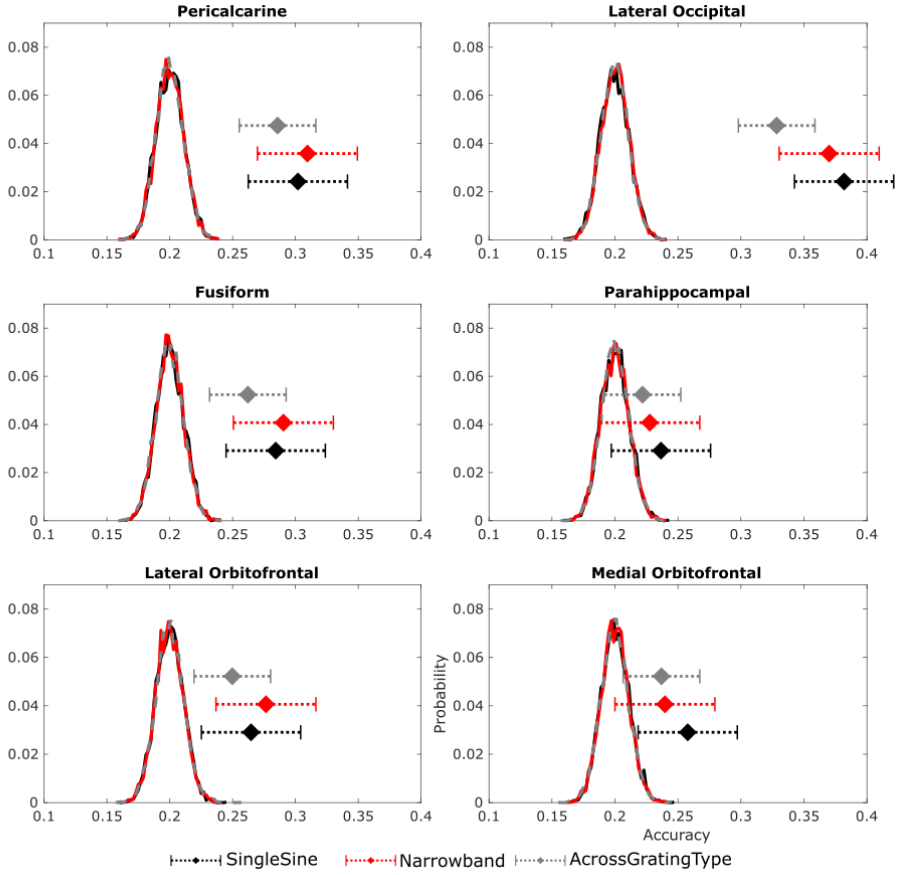


Figure 8: Spatial frequency classification per ROI. Rhombi indicate the average observed accuracy (see text for details). Classifiers are trained and cross-validated on single sinewave grating trials (black) or on narrowband grating trials (red). The gray elements show classifier generalization performance across grating type. Dotted margins illustrate standard errors of the mean. Distributions are the normalized permutation distribution (frequency of a given accuracy divided by number of permutations, 5000 permutations), i.e. empirical chance. Nominal chance level is 0.2 (5-class classification).

Discussion

We investigated how relative neural responses to single sinewave gratings and narrowband $1/f^2$ noise patches, centered on the same spatial frequency and orientation, evolve along the ventral visual pathway up to orbitofrontal visually responsive areas. We hypothesized that because the narrowband patches share a fundamental property with ecologically relevant natural scene images, namely the

$1/f^{\alpha}$ amplitude spectrum, the corresponding evoked neural signal would carry further along the ventral visual pathway than responses evoked by single sinewave gratings. To investigate our hypothesis, we parcellated individual participants' source reconstructions into atlas-based regions of interest, ordered according to their position in the visual hierarchy following standard models of vision (Van Essen, Anderson, & Felleman, 1992; Van Essen & Maunsell, 1983) and analyzed responses for each ROI.

We found no evidence that responses to narrowband compared to single sinewave gratings are more sustained over cortical space. Instead, narrowband gratings evoked generally larger response amplitudes when comparing sensor space responses over occipital channels, as well as marginally larger responses in hierarchically early ROIs such as pericalcarine and lateral occipital cortex.

Above-chance classifier accuracy suggests that it is possible to differentiate between the neural responses to stimuli with different spatial frequencies in all ROIs, for both single sinewave and narrowband gratings. This indicates that both early and high-level visual areas are sensitive to the spatial frequency of a grating, irrespective of any category information or statistical similarity to the regions preferred category in the stimulus. Classification could generalize between grating types in all ROIs, suggesting that spatial frequency coding in those ROIs is at least to some extent invariant to stimulus bandwidth. However, we also found classification performance to be higher when training and testing data belonging to the same rather than different grating type conditions. This indicates that SF modulated neural responses are not fully independent from stimulus bandwidth.

A potential criticism to the interpretation that higher level visual areas are sensitive to the SF content of stimuli is that the spatial resolution of our source reconstruction might be too low to resolve the activities of individual areas and, accordingly, SF tuning in these regions might erroneously reflect SF processing in early visual cortex. While beamformer techniques have been demonstrated to provide sufficient spatial specificity to differentiate between five retinotopic locations within V1 (Brookes et al., 2010) the spatial extent of sources produced by the alternative procedure based on minimum norm estimates (MNE) that we used here is considerably higher (Lin, Witzel, Zeffiro, & Belliveau, 2008). It is therefore possible that activations attributed to one ROI were contaminated by activations from other ROIs, a phenomenon known as signal leakage (Brookes, Woolrich, & Barnes, 2012). There are two reasons why this is unlikely to account for the present findings. Firstly, the inferior spatial specificity of MNE compared to beamformer techniques is not absolute but dependent on the signal SNR. Lin et al. (2006) found that at high SNR, such as the one provided by the SSVEP paradigm (Norcia et al., 2015), the point spread functions of source activations derived with beamformer and MNE techniques are more similar. The authors showed that both methods produce equally focal sources at high SNRs and even found an advantage for MNE

at very high SNRs (around 50). Secondly, we found spatial frequency specific responses in orbitofrontal cortex, a region with a large geodesic distance to the pericalcarine ROI (Oligschläger et al., 2017) in which the highest response to our stimuli was found. Even if we assume a large point spread function of 10mm full width at half maximum, it is unlikely that the effects observed in orbitofrontal ROIs purely reflect leakage from the calcarine ROI. Therefore, the high SNR and large distance between some of the ROIs in our study make it unlikely that low source resolution explains the SF tuning effects we observed across ROIs.

Another concern is that spatial frequency classification is driven by univariate differences in condition mean response strength rather than the spatial pattern associated with each SF. To mitigate this problem, we removed the mean activation from each ROI and condition prior to classification. However, while demeaning limits the direct effect of activation strength on classification, it does not equalize SNR across conditions. Therefore, the increased consistency of the higher SNR data could still contribute to classification performance.

Finally, higher level areas might passively inherit their higher mean activation in response to some SFs over others from early visual cortex. However, we are reasonably confident that SF classification was largely independent of differences in mean activation as narrowband and single sinewave-based classification performance does not mirror mean activation differences (compare e.g. pericalcarine and lateral occipital ROIs in figures 6 and 8).

A general concern in our study in the low number of participants and the corresponding deficiency in statistical power. To strengthen the confidence in the findings described here, more data is required. Further, the single trial estimates used in the source-space classifications were derived by multiplying the individual trial data with spatial filters based on the noise covariance over all conditions and repetitions. The parameters chosen here for noise covariance scaling (lambda = 3) might not have been optimal.

In conclusion, we found no evidence for our hypothesis that higher-level visual areas respond preferentially to narrowband gratings that follow a $1/f^2$ amplitude spectrum. We do however find that those regions are sensitive to the spatial frequency of a grating, and that the spatial patterns of activation we observe in our data are similar across grating type.

References

Aguirre, G. K., Zarahn, E., & D'esposito, M. (1998). An area within human ventral cortex sensitive to "building" stimuli: evidence and implications. *Neuron*, *21*(2), 373–383.

Andrews, B. W., & Pollen, D. A. (1979). Relationship between spatial frequency selectivity and receptive field profile of simple cells. *The Journal of Physiology*, *287*(1), 163–176.

Andrews, T. J., Clarke, A., Pell, P., & Hartley, T. (2010). Selectivity for low-level features of objects in the human ventral stream. *Neuroimage*, *49*(1), 703–711.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433–436.

Brookes, M. J., Woolrich, M. W., & Barnes, G. R. (2012). Measuring functional connectivity in MEG: a multivariate approach insensitive to linear source leakage. *Neuroimage*, *63*(2), 910–920.

Brookes, M. J., Zumer, J. M., Stevenson, C. M., Hale, J. R., Barnes, G. R., Vrba, J., & Morris, P. G. (2010). Investigating spatial specificity and data averaging in MEG. *Neuroimage*, *49*(1), 525–538.

Burton, G. J., & Moorhead, I. R. (1987). Color and spatial structure in natural scenes. *Applied Optics*, *26*(1), 157–170.

De Valois, K. K., De Valois, R. L., & Yund, E. W. (1979). Responses of striate cortex cells to grating and checkerboard patterns. *The Journal of Physiology*, *291*(1), 483–505.

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., … Hyman, B. T. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage*, *31*(3), 968–980.

Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Josa A*, *4*(12), 2379–2394.

Georgeson, M. A., & Sullivan, G. D. (1975). Contrast constancy: deblurring in human vision by spatial frequency channels. *The Journal of Physiology*, *252*(3), 627–656.

Goffaux, V., Duecker, F., Hausfeld, L., Schiltz, C., & Goebel, R. (2016). Horizontal tuning for faces originates in high-level Fusiform Face Area. *Neuropsychologia*, *81*, 1–11.

Goffaux, V., Peters, J., Haubrechts, J., Schiltz, C., Jansma, B., & Goebel, R. (2010). From coarse to fine? Spatial and temporal dynamics of cortical face processing. *Cerebral Cortex*, *21*(2), 467–476.

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., … Hämäläinen, M. S. (2014). MNE software for processing MEG and EEG data. *Neuroimage*, *86*, 446–460.

Gramfort, A., Papadopoulo, T., Olivi, E., & Clerc, M. (2010). OpenMEEG: opensource software for quasistatic bioelectromagnetics. *Biomedical Engineering Online*, *9*(1), 45.

Grill-Spector, K. (2003). The neural basis of object perception. *Current Opinion in Neurobiology*, *13*(2), 159–166.

Hansen, B. C., & Essock, E. A. (2004). A horizontal bias in human visual processing of orientation and its correspondence to the structural components of natural scenes. *Journal of Vision*, *4*(12), 5.

Haun, A. M., & Essock, E. A. (2010). Contrast sensitivity for oriented patterns in 1/f noise: Contrast response and the horizontal effect. *Journal of Vision*, *10*(10), 1.

Ishai, A., Ungerleider, L. G., Martin, A., Schouten, J. L., & Haxby, J. V. (1999). Distributed representation of objects in the human ventral visual pathway. *Proceedings of the National Academy of Sciences*, *96*(16), 9379–9384.

Kauffmann, L., Ramanoël, S., Guyader, N., Chauvin, A., & Peyrin, C. (2015). Spatial frequency processing in scene-selective cortical regions. *NeuroImage*, *112*, 86–95.

Korth, M., Rix, R., & Sembritzki, O. (1985). Spatial contrast transfer functions of the pattern-evoked electroretinogram. *Investigative Ophthalmology & Visual Science*, *26*(3), 303–308.

Kybic, J., Clerc, M., Abboud, T., Faugeras, O., Keriven, R., & Papadopoulo, T. (2005). A common formalism for the integral formulations of the forward EEG problem. *IEEE Transactions on Medical Imaging*, *24*(1), 12–28.

Leske, S., & Dalal, S. S. (2019). Reducing power line noise in EEG and MEG data via spectrum interpolation. *NeuroImage*, *189*, 763–776.

Lin, F.-H., Witzel, T., Ahlfors, S. P., Stufflebeam, S. M., Belliveau, J. W., & Hämäläinen, M. S. (2006). Assessing and improving the spatial accuracy in MEG source localization by depth-weighted minimum-norm estimates.

Neuroimage, *31*(1), 160–171.

Lin, F.-H., Witzel, T., Zeffiro, T. A., & Belliveau, J. W. (2008). Linear constraint minimum variance beamformer functional magnetic resonance inverse imaging. *Neuroimage*, *43*(2), 297–311.

Marcus, D., Harwell, J., Olsen, T., Hodge, M., Glasser, M., Prior, F., … Van Essen, D. (2011). Informatics and data mining tools and strategies for the human connectome project. *Frontiers in Neuroinformatics*, *5*, 4.

Murata, A., Fadiga, L., Fogassi, L., Gallese, V., Raos, V., & Rizzolatti, G. (1997). Object representation in the ventral premotor cortex (area F5) of the monkey. *Journal of Neurophysiology*, *78*(4), 2226–2230.

Nasr, S., & Tootell, R. B. H. (2012). A cardinal orientation bias in scene-selective visual cortex. *Journal of Neuroscience*, *32*(43), 14921–14926.

Norcia, A. M., Appelbaum, L. G., Ales, J. M., Cottereau, B. R., & Rossion, B. (2015). The steady-state visual evoked potential in vision research: a review. *Journal of Vision*, *15*(6), 4.

Norcia, A. M., Tyler, C. W., Hamer, R. D., & Wesemann, W. (1989). Measurement of spatial contrast sensitivity with the swept contrast VEP. *Vision Research*, *29*(5), 627–637.

Oligschläger, S., Huntenburg, J. M., Golchert, J., Lauckner, M. E., Bonnen, T., & Margulies, D. S. (2017). Gradients of connectivity distance are anchored in primary cortex. *Brain Structure and Function*, *222*(5), 2173–2182.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*, 1.

Peelen, M. V, & Downing, P. E. (2017). Category selectivity in human visual cortex: Beyond visual object recognition. *Neuropsychologia*, *105*, 177–183.

Penny, W. D., Friston, K. J., Ashburner, J. T., Kiebel, S. J., & Nichols, T. E. (2011). *Statistical parametric mapping: the analysis of functional brain images*. Academic press.

Pietrini, P., Furey, M. L., Ricciardi, E., Gobbini, M. I., Wu, W.-H. C., Cohen, L., … Haxby, J. V. (2004). Beyond sensory images: Object-based representation in the human ventral pathway. *Proceedings of the National Academy of Sciences*, *101*(15), 5658–5663.

Plichta, M. M., Heinzel, S., Ehlis, A.-C., Pauli, P., & Fallgatter, A. J. (2007). Model-based analysis of rapid event-related functional near-infrared spectroscopy (NIRS) data: a parametric validation study. *Neuroimage*, *35*(2), 625–634.

Quek, G. L., Liu-Shuang, J., Goffaux, V., & Rossion, B. (2018). Ultra-coarse, single-glance human face detection in a dynamic visual stream. *NeuroImage*.

Rao, C. R. (1948). The utilization of multiple measurements in problems of biological classification. *Journal of the Royal Statistical Society. Series B (Methodological)*, *10*(2), 159–203.

Rossion, B., Hanseeuw, B., & Dricot, L. (2012). Defining face perception areas in the human brain: a large-scale factorial fMRI face localizer analysis. *Brain and Cognition*, *79*(2), 138–157.

Ruderman, D. L., & Bialek, W. (1994). Statistics of natural images: Scaling in the woods. In *Advances in neural information processing systems* (pp. 551–558).

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. *Proceedings of the National Academy of Sciences*, *105*(11), 4447–4452.

Tolhurst, D. J., Tadmor, Y., & Chao, T. (1992). Amplitude spectra of natural images. *Ophthalmic and Physiological Optics*, *12*(2), 229–232.

Tootell, R. B. H., Hadjikhani, N. K., Vanduffel, W., Liu, A. K., Mendola, J. D., Sereno, M. I., & Dale, A. M. (1998). Functional analysis of primary visual cortex (V1) in humans. *Proceedings of the National Academy of Sciences*, *95*(3), 811–817.

Van der Schaaf, van A., & van Hateren, J. H. van. (1996). Modelling the power spectra of natural images: statistics and information. *Vision Research*, *36*(17), 2759–2770.

Van Essen, D. C., Anderson, C. H., & Felleman, D. J. (1992). Information processing in the primate visual system: an integrated systems perspective. *Science*, *255*(5043), 419–423.

Van Essen, D. C., & Maunsell, J. H. R. (1983). Hierarchical organization and functional streams in the visual cortex. *Trends in Neurosciences*, *6*, 370–375.

Watson, D. M., Hartley, T., & Andrews, T. J. (2014). Patterns of response to visual scenes are linked to the low-level properties of the image. *NeuroImage*, *99*, 402–410.

Chapter 3: Coarse-to-fine information integration in human vision

Corresponding publication:
Petras, K., ten Oever, S., Jacobs, C., & Goffaux, V. (2019). Coarse-to-fine information integration in human vision. NeuroImage, 186, 103-112.

Kirsten Petras[1*], Sanne ten Oever[2], Christianne Jacobs[1], Valerie Goffaux[1,2,3]
(1) Research Institute for Psychological Science, Université Catholique de Louvain, Louvain-la-Neuve, Belgium
(2) Department of Cognitive Neuroscience, Maastricht University, Maastricht, The Netherlands
(3) Institute of Neuroscience, Université Catholique de Louvain, Louvain-la-Neuve, Belgium
(*) corresponding author. E-Mail address: *kirsten.petras@uclouvain.be*

Summary

Coarse-to-fine theories of vision propose that the coarse information carried by the low spatial frequencies (LSF) of visual input guides the integration of finer, high spatial frequencies (HSF). Whether and how LSF modulates HSF processing in naturalistic broad-band stimuli is still unclear. Here we used multivariate decoding of EEG signals to separate the respective contribution of LSF and HSF to the neural response evoked by broad-band images. Participants viewed images of human faces, monkey faces and phase-scrambled versions that were either broad-band or filtered to contain LSF or HSF. We trained classifiers on patterns evoked by filtered scrambled stimuli and generalized decoding to broad-band scrambled and intact trials. We found reduced HSF contribution when LSF was informative towards image content, indicating that coarse information does guide the processing of fine detail, in line with coarse-to-fine theories. We discuss the potential cortical mechanisms underlying such coarse-to-fine feedback.

Introduction

Naturalistic visual input contains a wide range of spatial frequencies (SF) that are closely correlated in space and that are processed in an integrative fashion to support visual perception. Several theories have proposed that the visual system integrates visual input in a coarse-to-fine (CtF) manner. In this framework, coarse, low spatial frequency (LSF) information is processed first and quickly projects from primary visual cortex to higher level visual areas, which generate a feedback signal that subsequently guides the processing of finer-grained high spatial frequency (HSF) information (Marr, 1982; Watt, 1987; Schyns and Oliva, 1994; Bullier, 2001; Bar, 2003, 2004; Bar et al., 2006; Hegdé, 2008). For example, when you look at a dog, its general shape, carried by LSF, helps with the integration of finer details carried by HSF, such as the individual hairs of the dog's fur. Using LSF information to guide HSF processing is feasible, because both ranges are strongly correlated in natural images (i.e. the shape of a dog restricts where the fur can be). CtF strategies have been shown to be highly beneficial in terms of both speed and efficiency in many computer vision applications (e.g. Burt, 1988; Gavrila and Philomin, 1999; Zhou et al., 2013; Zhang et al., 2014), as well as in computational models of higher level vision (Kay and Yeatman, 2017). But despite an abundance of evidence for coarse over fine *precedence* in response to simple (see e.g. Jones and Keck, 1978; Parker, 1980; Musselwhite and Jeffreys, 1985; Parker and Dutch, 1987; Watt, 1987; Hughes, Nozawa and Kitterle, 1996; Mihaylova, Stomonyakov and Vassilev, 1999; Mazer et al., 2002 for dots, lines and gratings), as well as to more complex stimuli like faces and natural scenes (Parker, Lishman and Hughes, 1992, 1997; Schyns and Oliva, 1994; Halit et al., 2006; Peyrin et al., 2006; Vlamings, Goffaux and Kemner, 2009; Goffaux et al., 2010; Lu et al., 2018), direct evidence for coarse-to-fine *integration* is still scarce. Coarse to fine integration requires not only that coarse image information is processed quicker than finer details, but also that this initial processing influences later stages of HSF information integration. To be beneficial, CtF integration should reduce either the required processing time and/or the metabolic cost of visual processing by alleviating resources from processing redundant detail information. In order to experimentally investigate CtF integration, several fundamental characteristics of natural visual input need to be mimicked. A first characteristic is that natural images are broad-band, i.e. they contain a combination of both low and high spatial frequencies. Second, the different ranges of SF composing an image are highly correlated in space: without such correlation (i.e., if LSF offers no information on the content of HSF), CtF integration is not computationally advantageous. Finally, natural images follow a roughly $1/SF^{\alpha}$ amplitude spectrum: the LSF components of a natural image generally have much higher contrast than its HSF components.

In spite of the broad-band nature of visual input, previous studies addressed CtF processing mostly by comparing the neural responses to narrow-band spatial frequency filtered stimuli (e.g. Goffaux et al., 2010; De Cesarei and Codispoti, 2013; Musel et al., 2014). While they provide a solid foundation for coarse over fine *precedence*, they cannot address CtF *integration*, because SF ranges were presented separately in space, time, or both. Other studies have used so-called hybrid stimuli, combining LSF and HSF from different images, for example the LSF components of a female face with the HSF components of a male face (e.g. Schyns and Oliva, 1999). While hybrid stimuli present an observer with high and low frequencies simultaneously, the uncorrelated nature of LSF and HSF is in violation of the second characteristic of natural images, namely spatially correlated LSF and HSF signals. Additionally, the LSF and HSF stimuli used in past studies had different contrasts, which causes a considerable confound, because visual sensitivity and stimulus contrast are closely linked (Robson, 1966). Other studies equalized stimuli for contrast across SF conditions (Goffaux and Rossion, 2006; Peyrin et al., 2006; Vlamings, Goffaux and Kemner, 2009; Mu and Li, 2013) which is in violation of the $1/SF^{\alpha}$ amplitude spectrum of natural images and might bias visual analysis strategies (Kauffmann et al., 2015). Notably, one approach to avoid the problems that arise through filtering and contrast matching is to mask out information in a given spatial frequency range by replacing the information with its phase-scrambled version. This strategy renders the masked frequency range uninformative without interfering with the natural amplitude spectrum (Näsänen, 1999; Ojanpää and Näsänen, 2003). However, while the resulting images meet the requirements of being broad-band and following a natural amplitude spectrum, the masking disrupts the correlation between LSF and HSF necessary for CtF integration to occur. In summary, by attempting to disentangle LSF and HSF at the level of the stimulus, all of the above research employed stimuli that departed considerably from the fundamental characteristics of naturalistic visual input which are essential to successful CtF integration.

In the present study, we employ a novel paradigm using multivariate pattern classification to tease apart LSF and HSF contributions during the processing of broad-band (BB) stimuli. Separating SF at the level of the cortical response rather than at the level of the stimulus allows us to preserve the fundamental properties of naturalistic broad-band vision, thus allowing to investigate CtF in more ecological settings than achieved so far. With Electroencephalography (EEG) we measure LSF- and HSF-related neural responses expressed as scalp topographies. Then, we train a set of linear classifiers to differentiate between these responses and test their generalization to broad-band visual stimulation over time. Namely, the classifier labels the broad-band response as either LSF or HSF, depending on which one is more prominent in the broad-band topographies; therefore, we interpret classifier prediction as *spatial frequency dominance*. Comparing SF dominance patterns over

time allows us to directly address the essential prediction of CtF theories: LSF processing influences subsequent HSF processing in broad-band visual input. We include four different stimulus types: human faces, monkey faces, and both their phase-scrambled versions. Human faces were chosen as stimuli because they elicit a strong and well characterized EEG response (Eimer, 2011). Monkey faces share many of the same features without being equally ubiquitous in the normal human visual diet. While in both categories of intact images LSF is informative with respect to HSF, this is not the case for the phase-scrambled ones. Further, while both human and monkey faces show similar low level image properties and rely on largely similar resources for their processing, human face processing has been shown to be exceptionally efficient compared to other visual categories (Farah et al., 1998; Haxby, Hoffman and Gobbini, 2000). This effect is thought to strongly rely on LSF information (Sergent, 1982, 1986; Fiorentini, Maffei and Sandini, 1983; Goffaux and Rossion, 2006; Goffaux et al., 2010; Richler, Cheung and Gauthier, 2011) . Thus, responses to human and monkey faces should be similar if basic image properties drive coarse-to-fine integration but differ if higher level content information, which presumably is more readily available for human faces, is used instead. Consequently, we expected that when LSF is informative towards HSF content (i.e., in intact stimulation conditions), early LSF dominance would coincide with a reduced need to process the correlated HSF information, signified by reduced HSF dominance later in processing time (i.e. CtF integration) and that this effect would be stronger for human compared to monkey faces. To avoid any influence of semantic image content, we trained all classifiers exclusively on the phase-scrambled images.

Methods

Participants

21 healthy volunteers (mean age 26.7, range 21-34, 10 female) with normal or corrected to normal visual acuity participated in the EEG experiment. Data were collected as part of a technical development project aimed at improving EEG source reconstruction. Therefore, a subset of the participants also took part in a separate fMRI experiment not further described here and all EEG recordings were performed while participants were lying on their back in a mock-scanner. All participants gave written informed consent and were reimbursed for their participation in the form of gift vouchers (7.5/hour). The EEG experiment lasted 55 minutes with an additional set-up time of 15-35 minutes. The subjects who also participated in the MRI experiment additionally underwent a 10-minute anatomical scan and two functional localizer runs of 12 minutes each. All procedures have been approved by

the ethical committee of the Université Catholique de Louvain (approval number: 2016/13SEP/393).

Stimuli

Stimuli were created from seven original frontal views each of female human and monkey faces. Images were scaled and aligned so that the distance between eyes and mouth was equal for all images, cropped to the same oval occlusion and pasted onto a uniformly gray background. They were then converted to grayscale and luminance and root-mean-square (RMS) contrast was equalized using custom made functions in Matlab (Mathworks, 2016). Images of faces usually follow a roughly $1/SF^{\alpha}$ amplitude spectrum, were α is close to 2, giving considerably higher spectral energy to low than to high spatial frequencies. Since visual processing is strongly influenced by the contrast of the stimulus (De Valois and De Valois, 1980; Boynton et al., 1996), it is common practice to equalize RMS contrast between stimuli by enhancing contrast in HSF images after filtering (Vlamings, Goffaux and Kemner, 2009; Goffaux et al., 2010; Mu and Li, 2013; Kauffmann et al., 2015). However, since such a manipulation would result in non-natural amplitude spectra we developed the following normalization to avoid contrast differences between LSF and HSF stimuli while maintaining a natural spectrum for the composite broad-band (BB) images (see figure 1). We first constructed 5^{th} order Butterworth frequency filters with cut-offs chosen such that, averaged over all images, the summed Fourier-space amplitudes of LSF and HSF images were equal. Since most energy is contained in extremely low SF, our cutoff manipulation would lead to LSF images containing only very coarse visual information if applied to full spectrum images. To prevent such large imbalances in information content, we excluded the first 5 cycles per face (cpf) prior to determining the filter cut-offs. The Fourier amplitude of each image was then multiplied with the resulting filters and transformed back into image space. To create BB images, the Fourier-amplitudes of the original images were multiplied with the sum of the low (5-8 cpf) and the high (>8cpf) SF filters and then also transformed back into image space. Importantly, there was no post-hoc contrast enhancement of HSF images. Instead, the design of the Butterworth-filters ensured that mean contrast was 'naturally' similar between HSF and LSF images, namely 0.0185 and 0.018 respectively (see figure 1).

To derive meaningless noise images with power spectra similar to those of the original images, we created phase-scrambled versions of full spectrum images by replacing the phase information of the Fourier-Transform with random noise. Operations in Fourier space require input images to be rectangular and cannot be restricted to the region occupied by the face.

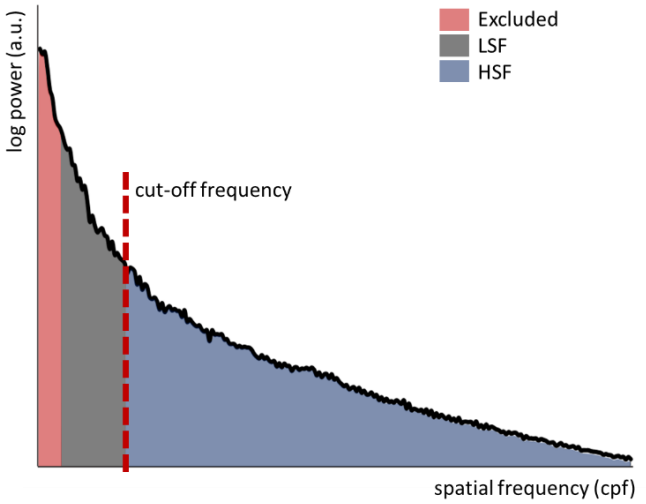


Figure 1. The cut-off frequency for the Butterworth filters was determined as the frequency for which on average over all images, the area under the curve of the Fourier amplitude spectrum was equal to either side of the cut-off. To avoid very small bandwidths in the LSF condition, we excluded the first 5 cycles per image before determining the cut-off. Note that this is a conceptual illustration and not to scale.

Therefore, the phase-scrambling may artificially increase LSF energy due to the low frequency uniform background leaking into the oval face area and vice versa. To minimize the effects of such leakage, we employed the following iterative approach to phase scrambling: Images were cropped to the square that most closely matched the occlusion and phase-scrambled in Fourier space before being transformed back into image space. The original, non-scrambled face pixels were pasted onto the resulting phase-scrambled images. This procedure was repeated 100 times so that face areas and backgrounds had similar power spectra before moving on to the next step. The scrambled face areas were then pasted back onto the original gray backgrounds. In effect, by making the spectral properties of face and background pixels more similar, this procedure minimized contaminations of the scrambled image by the uniform background (see figure 2 B). The images were then subjected to the same Butterworth filter procedure as the phase-intact images to derive LSF, HSF and BB versions. Images were displayed on a standard LCD monitor (1900x1200px) with an absolute size of the face area of 5.8cm (max. height of the oval shape) x 4.7cm (max. width of the oval shape) at a viewing distance of approx. 60 cm thus covering ~5.5° x ~4.5° of visual angle.

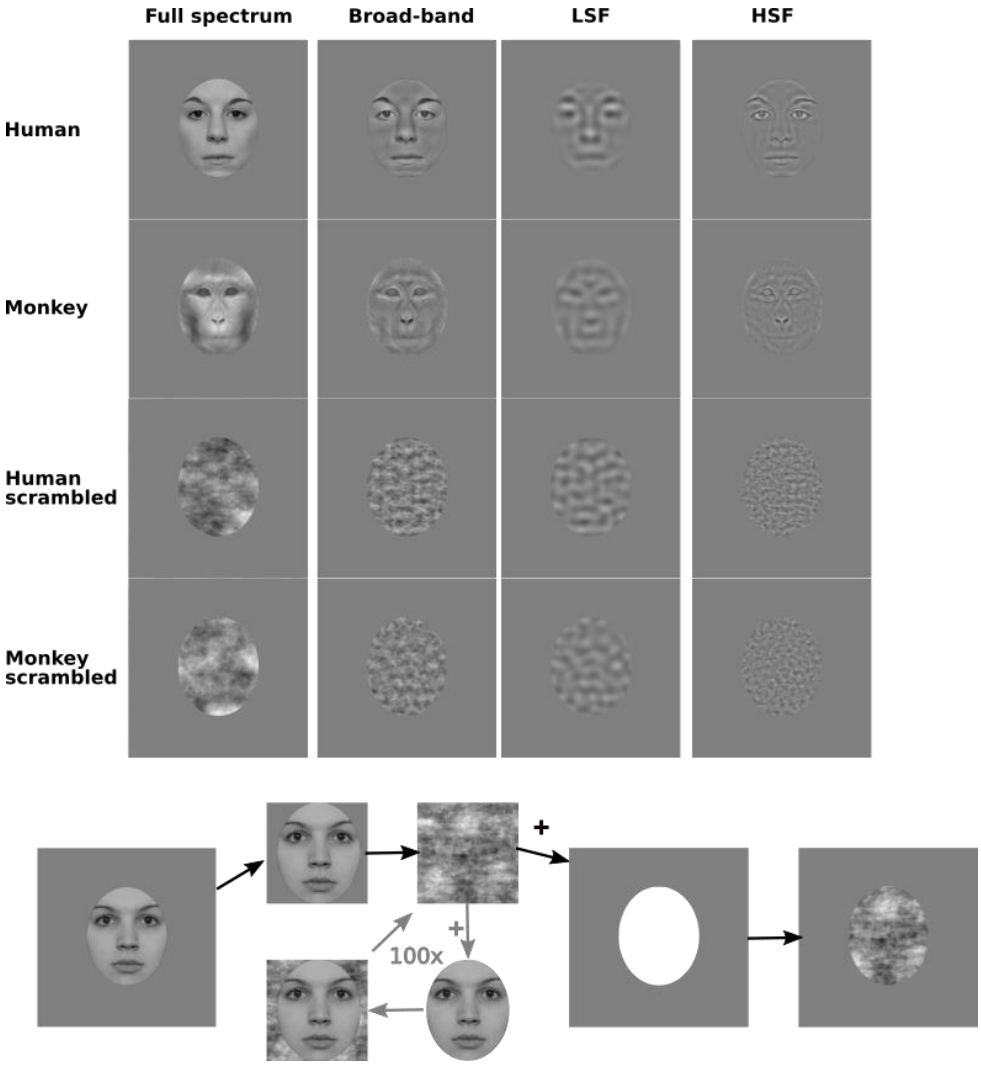


Figure 2. Upper panel: Example stimuli. Full spectrum grayscale pictures of human and monkey faces (leftmost column, note that the original images were never shown to the participant) were filtered to contain low and high, only low, or only high spatial frequencies. All faces were aligned at the horizontal connecting the eyes and scaled such that the vertical distance between eyes and mouth was equal across images. Lower panel: Images with similar amplitude and orientation spectra as the original face images, but without any semantic content were created. Face images on gray background were cropped to the closest square surrounding the oval face outline and scrambled in Fourier-phase. To reduce the loss of spatial frequency information due the uniform background leaking into the oval image area, the intact face was pasted back onto the scrambled image and the whole process was repeated 100 times before adding the original gray background.

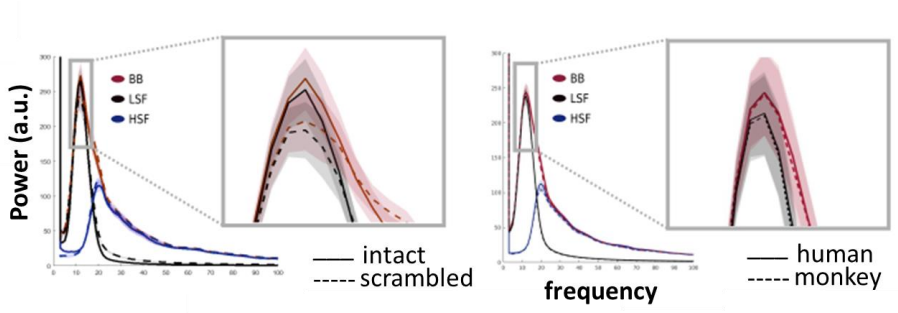


Figure 3. Left: Power spectrum comparing intact and scrambled images. Spectra were averaged over human and monkey images. Right power spectra comparing human and monkey images. Note that differences between categories are minimal and well within standard error margins.

Procedure

During the experiment, participants were passively viewing intact and scrambled images of human faces and monkey faces in the 3 filter conditions (LSF, HSF and BB). Intact images and scrambled images were presented in blocks. Each stimulus was on screen for 300ms with a dark grey fixation cross superimposed in the center of the image. Block order was counterbalanced between participants. Stimulus order within each block as well as inter-trial-interval (1.5-2.5 sec) were counterbalanced via an m-sequence (Buračas and Boynton, 2002) to avoid any potential order effects. Participants underwent a total of 1440 trials subdivided into 6 blocks. Each block contained 206 experimental trials and 34 catch trials in which participants had to indicate whether a black shape appearing on the current image was a circle or a square by pressing one of two buttons on a button box. Catch trials, as well as trials directly following a catch trial were removed from further EEG analysis.

EEG recording and analysis

An EEG geodesic sensor net (Electrical Geodesics Inc., Eugene, OR, USA) was soaked in a potassium chloride solution for 10 minutes to facilitate conductivity of the sponges placed within plastic cups between scalp and electrodes. The net was fitted on the participant's head such that electrode Cz, which served as a reference for the remaining 255 electrodes, was located on the intersection of the midlines between nasion and inion and the preauricular points. We then carefully removed hair trapped between electrodes and scalp and re-moistened sponges that did not meet our scalp-electrode impedance requirements (<40kΩ at the beginning of the recording). EEG signals were amplified with a high input-impedance of ~200MΩ

using a Net-Amps dense array amplifier (Electrical Geodesics Inc., Eugene, OR, USA) and sampled at a rate of 1000Hz. To prevent sponges from drying out during the long recording session, we wrapped the participants head in cling film held together with tape which also served to keep electrodes in place and close to the scalp once participants laid on their backs. We could thus maintain scalp-electrode impedances < 50kΩ throughout the recordings for almost all electrodes. Participants were placed in a mock-MRI-scanner (no magnetic field present) and viewed a screen at a distance of 60cm via a mirror mounted to the (not connected) head coil.

EEG data were preprocessed using the Fieldtrip toolbox (Oostenveld et al., 2011) under Matlab. Data were re-referenced to the average of all electrodes and band pass filtered to 0.5 to 40 Hz. Trials were segmented into epochs ranging from -500ms to 1500ms around stimulus onset and down-sampled to 256Hz. We then manually removed trials with excessive muscle artifacts (but not blinks) and subjected the remainder of the data to an ICA algorithm (Runica, implemented in Fieldtrip) and subsequently removed components containing ocular artifacts. Finally, to reduce the number of features and as a consequence the computational demands for classification, only a region of interest consisting of the 47 electrodes covering occipital and temporal regions were a-priori selected (see figure 4 for locations).

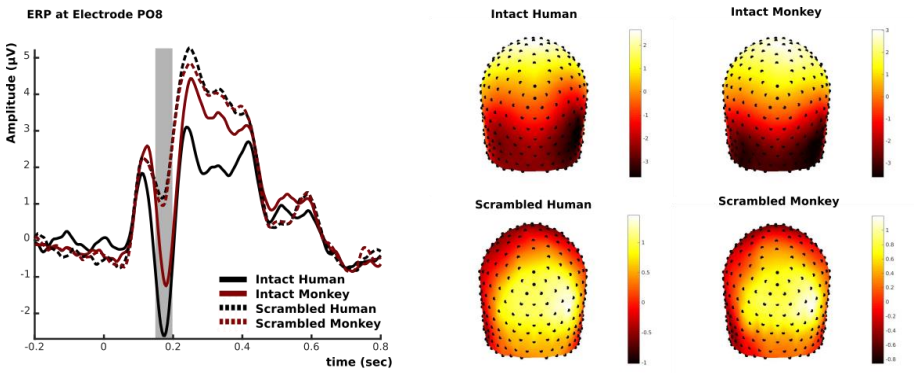


Figure 4. Left: Event related potential at electrode PO8. Note the typical N170 response (grey shaded area) that is most prominent for human faces and least prominent for scrambled images. Right: Averaged EEG scalp topographies at the latency of the N170 component (160-180ms). Note that both human faces and monkey faces evoke a right lateralized negativity over occipito-temporal areas.

Classification

The aim of the study was to investigate the contribution of low and high SF over the time course of broad-band image-processing. To avoid any influence of semantic content on our classifiers, we only used phase-scrambled image trials for training. Classifiers were trained to discriminate between LSF and HSF scrambled trials. We then used this "low vs high scrambled" model and evaluated its prediction in response to broad-band scrambled and intact trials. This provided us with information about the dominance of either low or high spatial frequency in the EEG signal during broad-band viewing conditions.

We used the support vector machine algorithm provided by libsvm (Chang and Lin, 2011) running under Matlab for all classifications. We employed a linear kernel and kept the regularization parameter c equal to 1 throughout classifications. Epochs included the -100 to 600ms time window around stimulus onset (rounded to fit sampling rate). Then, we created time bins by shifting a 12ms window in 8ms steps (rounded to fit sampling rate) resulting in 90 bins. For each time bin we averaged amplitudes over the 4 samples per channel contained in it and, separately for all conditions, transformed feature-vectors into unit vectors such that:

$\frac{u^{=u}}{|u|}$ where |u| is the norm of the original vector u given by

$$|u| = \sqrt{\sum_{n=1}^{n} x_n^2}$$

To validate our LSF vs. HSF model we pseudo-randomly split the training data for each subject into training and validation sets retaining 20% of the original data for model validation. This procedure was repeated 150 times to ensure the greatest variability in fold composition that was computationally feasible. Statistical significance of above chance classification was assessed for each data point via two-sided paired samples t-tests against the empirical chance level of each participant. Data across all participants was corrected for multiple comparisons using cluster based permutation testing (Oostenveld et al., 2011) at a cluster alpha of 0.05 using the maximum sum of all t values within a cluster as a test statistic and testing against the Monte-Carlo estimates of the critical values from the permutation distribution (10 000 permutations). The empirical chance distribution was derived by label permutations. At each randomization, two models were generated and assessed: one with the original class labels and one with trial labels randomly permuted before classification performance was assessed. All reported accuracies for within condition classification are the average over all randomizations and participants. Please note that, because we kept class frequencies balanced in all

training and validation sets, the commonly used area under the curve (AUC) and plain accuracy are the same. We therefore report accuracy throughout the results section.
The generalization of the LSF/HSF scrambled classifier was assessed on all available trials for each BB condition (human, monkey, intact and scrambled) for each of the 150 models per time bin. To assess classifier generalization performance all BB trials were arbitrarily labeled as HSF. Off-chance generalization was interpreted as LSF or HSF dominance dependent on bias direction (<50% as LSF and >50% as HSF dominance). The statistical procedure (paired samples t-tests against empirical chance and cluster correction for multiple comparisons on the group level) was identical to the model validation procedure.
For between condition contrasts (Intact vs. Scrambled and Human vs. Monkey) we contrasted the individual condition cross-time matrices with each other and tested for statistical significance using paired sample t-tests. Results were again cluster corrected at a cluster alpha of 0.05.

Results

Successful spatial frequency decoding in scrambled images

The SVM classifiers were able to separate LSF and HSF trials significantly better than empirical chance starting from around 90ms after stimulus onset, [$p<0.001$, cluster-based permutation tests for multiple comparisons using the maximum sum of cluster t-values against empirical chance (see methods for details), cluster statistic = 5153.4] until about 560ms after stimulus onset (Figure 5). Although class frequencies were balanced in all training and testing folds, on average over all iterations, trials and subjects, classifiers predicted topographies as belonging to LSF-trials 51,2% and HSF-trials 48.8% of the time. When training time and testing time match (no generalization case) classification shows no bias (see suppl. figure 3). To assess the classification stability over time, we tested generalization of each classifier to every time-sample of the cross-validation trials (for procedure see King and Dehaene (2014) and methods). If the evoked scalp patterns used for classifier training and prediction remain similar across those time samples a classifier trained at a given time-sample t can still successfully predict SF class at a later (or earlier) time-sample t'. As expected, the highest classification accuracy was found around the diagonal of the time-generalization matrix; i.e. when training and testing times were identical (no generalization). While the performance of classifiers trained at early time windows (<200ms post stimulus onset) rapidly decayed with temporal distance between training and testing samples, classifiers trained on time windows later than 200ms post stimulus onset remained efficient in predicting SF condition for up to 200ms. This difference in temporal generalizability could indicate either a

stabilization of the cortical representation over time, or it could simply be due to a larger variance in the later evoked responses. All models (i.e. trained classifiers) resulting from the cross-validation procedure were stored to use for the subsequent analyses.
Notably, SF dominance is not only determined by the evolution of testing time (pattern changes in x direction of the temporal generalization matrix), but also by the training time, i.e. the model used to perform the prediction (patterns change along the y-direction, see figures 5 C & D and figure 6). This model evolution indicates that the classifier used different criteria to label a trial as LSF or HSF dependent on the progress of visual processing. Based on visual inspection we differentiated at least two different main time-courses of model validation. The first generalization pattern depends on models trained at times ranging from ~110ms to ~220ms after stimulus onset and the second on models trained at times ranging from ~230ms to ~440ms after stimulus onset. These latencies coincide with the P1-N1 and the P2-P3 ERP complex, respectively. We did not foresee this shift of critical information used by the classifier but speculate that over the course of visual processing different features, potentially originating from different cortical sites, are indicative of the spatial frequency content of the processed stimulus (see discussion). We however remain cautious to draw any strong conclusions based on post-hoc explanations.

Coarse-to-fine pattern of spatial frequency dominance in broad-band image processing

We applied the models trained to differentiate between LSF and HSF phase-scrambled stimuli to broad-band images of intact and phase-scrambled human and monkey faces (where LSF and HSF contributed equally to the overall image contrast) and assessed their predictions. Using the same temporal generalization approach as above, we applied each of the LSF vs. HSF classifiers to each of the BB trial time windows. Because of LSF and HSF components contribute equally to our BB images, classifier prediction now indicates the relative resemblance of the current scalp pattern to those associated with LSF and HSF trials, hereafter referred to as *SF dominance*, rather than SF class membership.
For all BB conditions, classifier prediction deviated significantly from chance only when 'training-time' fell into the window of successful classification between LSF and HSF trials, that is starting from 90ms after stimulus onset on (see figure 5 C-D). In other words, when a model was successfully trained to classify LSF vs. HSF scrambled images, this model generalized to predict BB trials as either LSF or HSF in a non-random fashion. These findings support the validity of our generalization results.

SF dominance patterns evoked by intact BB images (pooled over human and monkey faces) contain three post stimulus clusters of significant LSF dominance [128ms-207ms testing time, cluster statistic = 1206.6, $p<0.01$], [238ms-395ms testing time, cluster statistic = 728.8, $p<0.05$] and [451ms-600ms testing time, cluster statistic = 1147.3, $p<0.01$] and one cluster of significant HSF dominance [215ms-427ms testing time, cluster statistic = 589.1, $p<0.05$]. In contrast, the SF dominance matrices evoked by scrambled BB images contain only two post-stimulus clusters of significant LSF dominance [151ms-530ms testing time, cluster statistic = 831.7, $p<0.01$] and [458ms-600ms testing time, cluster statistic = 738.6, $p<0.01$] as well as two separate clusters of significant HSF dominance [200ms-466ms testing time, cluster statistic =1588.6, $p<0.01$] and [199ms-466ms testing time, cluster statistic = 762, $p<0.05$]. All latency ranges indicated here refer to testing time (x-axes of the temporal-generalization plots). Interestingly, we also find significant LSF dominance in the pre-stimulus period for both intact and scrambled BB conditions [Intact: -100ms - 90ms testing time, cluster statistic = 2523, $p<0.001$; Scrambled: -100ms – 100ms testing time, cluster statistic = 3188, $p<0.001$]. This pre-stimulus LSF dominance did not differ significantly between conditions (see figure 6 E). Because the stimulus is removed from the screen during the baseline period, what remains is the very low SF of the uniformly gray screen (0 cpi). Therefore, the pre-stimulus (empty) screen is sensibly labeled as LSF by the classifier (also see suppl. figure 3).

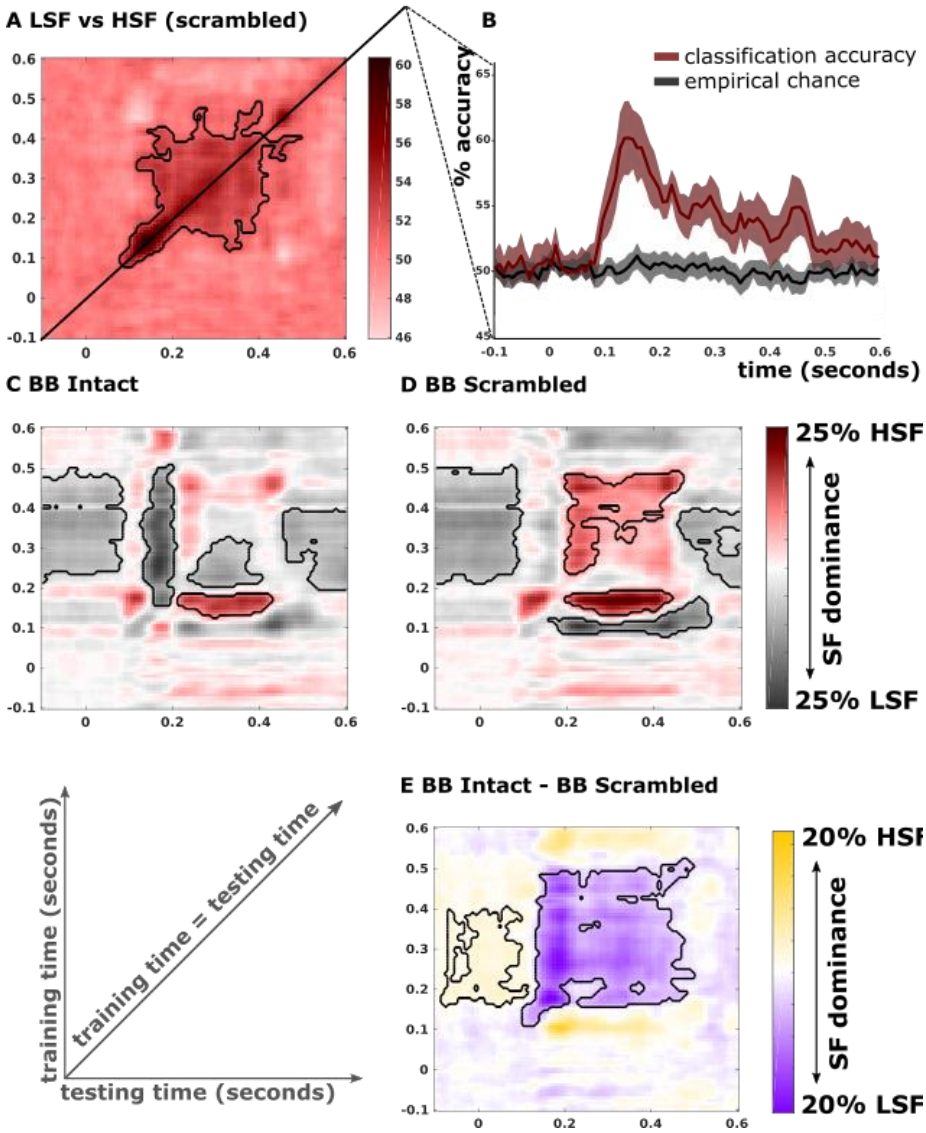


Figure 5. (A) Time generalized spatial frequency decoding between LSF and HSF scrambled images. The white contour indicates decoding performance significantly above empirical chance. (B) Time-course of spatial frequency decoding against empirical chance. Note that this is identical to the diagonal of the matrix in (A). Margins indicate standard error of the mean. (C) Spatial frequency dominance in intact broad-band images. Black contour indicates significant low or high spatial frequency dominance. (D) Spatial frequency dominance in scrambled broad-band images. (E) Difference between C and D. Black contour indicates significantly stronger low spatial frequency dominance in intact images.

Coarse-to-fine integration for intact, but not scrambled images

Low and high spatial frequencies are spatially aligned in intact images, satisfying the prerequisite for CtF integration. This is not the case for scrambled images. To statistically evaluate the differences in SF dominance between intact and scrambled BB conditions we contrasted predictions for BB-intact and BB-scrambled conditions. Results show significantly larger LSF dominance in intact compared to scrambled images [105ms - 482ms testing time cluster statistic = 6224.8, $p<0.001$ (see figure 5 E). Because we use binary classifiers, larger LSF dominance can equally be interpreted as reduced HSF dominance. However, the early LSF component is present only in the intact, but not in the scrambled trials whereas the late HSF component is present only in the scrambled, but not in the intact images. We therefore interpret the contrast as indicating an early LSF dominance and a late HSF reduction for intact, compared to scrambled images in line with CtF integration when low spatial frequencies are predictive of high spatial frequency information.

High level category dependence of spatial frequency dominance

The use of intact and phase-scrambled trials allowed us to compare between conditions in which low spatial frequency is or is not informative about high spatial frequency. Comparing human trials to monkey trials taps into a finer functional distinction. Indeed human and monkey faces are two visually similar classes of stimuli in which low spatial frequency is similarly informative towards high spatial frequencies in the image space. Images of human and monkey faces also share similar amplitude spectra (see figure 3) and cortical resources for their processing. However, the human visual system has been shown to be particularly sensitive to conspecific faces and highly efficient in their processing indicating further specialization for this stimulus class (Bruce and Young, 1986). Generalizing SF classifiers to BB images of human and monkey faces separately, we found EEG responses to human faces to be significantly LSF dominated in the post stimulus period [128ms-200ms testing time, cluster statistic = 1221.6, $p<0.01$], [246ms-400ms testing time, cluster statistic = 909.7, $p<0.05$] and [460ms-600ms testing time, cluster statistic = 1044.7, $p<0.01$]. While responses to monkey faces also showed early [144ms-207ms testing time, cluster statistic = 1146.8, $p<0.01$] and later [458ms – 600ms testing time, cluster statistic = 1133.4, $p<0.01$] LSF dominance, they additionally demonstrated an intermediate period of HSF dominance [215ms-419ms testing time, cluster statistic = 638.7, $p<0.05$].
Directly comparing SF dominance evoked by intact human faces to SF dominance evoked by intact monkey faces, we find significantly higher LSF dominance for the human face trials [96ms-175ms testing time, cluster statistic = 520.3, $p=0.01$; 246ms-420ms testing time, cluster statistic = 667.6, $p<0.01$]. Again, our use of binary classifiers limits the interpretation of directionality of the effect. Only in

combination with the generalization patterns against empirical chance does it become apparent, that the stronger LSF dominance later in trial-time for human face trials is in fact driven by the absence of the HSF dominance period present in monkey trials. Importantly, this difference in SF dominance patterns is not driven by purely low-level image properties: results show strikingly similar patterns for both categories when classification is based on their scrambled versions (see figure 6 F). Both human and monkey phase-scrambled images evoked LSF as well as HSF dominant patterns in the post stimulus period [Human scrambled LSF: 159ms-529ms testing time, cluster statistic = 946.6, $p<0.05$; 458ms – 600ms testing time, cluster statistic = 867.9, $p<0.05$]; [Human scrambled HSF: 199ms – 458ms testing time, cluster statistic = 3134.2, $p<0.001$; 207ms – 443ms testing time, cluster statistic = 946.6, $p<0.05$] and [Monkey scrambled LSF: 183ms – 521ms testing time, cluster statistic = 760.5, $p<0.05$]; [Monkey scrambled HSF: 199ms – 466ms testing time, cluster statistic = 1486.3, $p<0.01$; 199ms – 435ms testing time, cluster statistic = 812.3, $p<0.05$]. We found no significant differences between the two categories of phase-scrambled stimuli ($Ps>0.05$).

We therefore conclude, that intact, more than scrambled, and human, more than monkey, face images evoke early LSF dominant responses. Furthermore, the late HSF dominance is significantly reduced for intact and particularly for human face trials. These findings are in line with CtF theories that predict a LSF precedence along with a reduction of HSF processing load when LSF is predictive of HSF content.

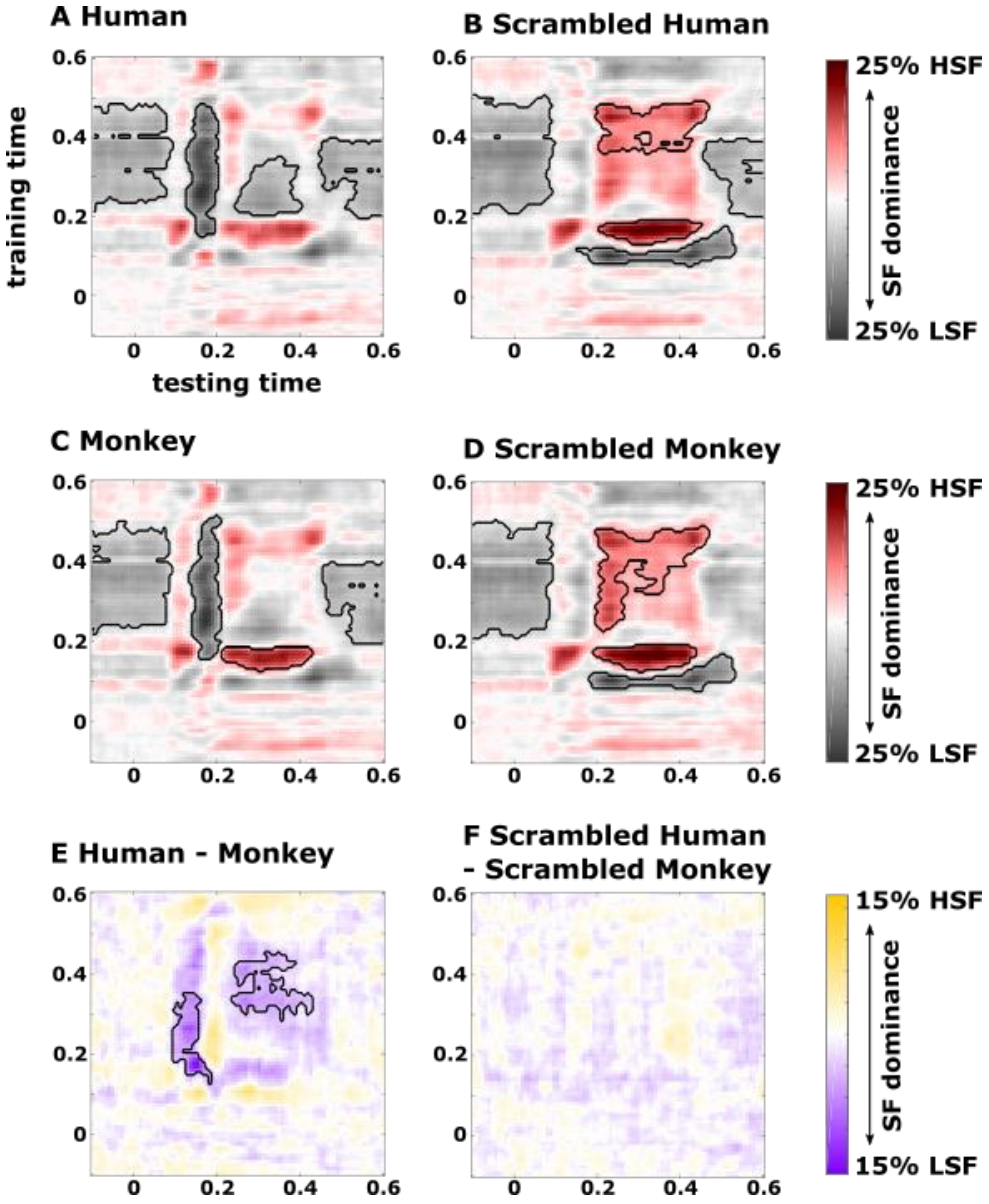


Figure 6. (A) and (B): Spatial frequency dominance in intact and scrambled human face images. Black contour indicates significant low spatial frequency dominance. White contour indicates significant high spatial frequency dominance. (C) and (D): Spatial frequency dominance in intact and scrambled monkey face images. (E) Difference between (A) and (C). Black contour indicates significantly stronger low spatial frequency dominance in human, compared to monkey images. (F) Difference between (B) and (D). No differences were found in spatial frequency dominance between scrambled human and scrambled monkey faces.

Discussion

Coarse-to-fine theories are a highly influential alternative to feedforward models of visual processing. So far, the main empirical support for CtF processing came from studies presenting LSF or HSF in isolation and showing coarse over fine temporal precedence. Here we directly investigated the CtF integration underlying the processing of more naturalistic broad-band input. In intact images, i.e. when image LSF were aligned to, and therefore informative about, HSF content, we found a reduction in the late contribution of HSF to the neural responses to BB, suggesting that informative LSF does modulate the processing of HSF information. Further, we found this pattern to be more prominent in response to human faces, for which the human visual system has developed robust perceptual templates and therefore strong predictions compared to monkey faces. These results are consistent with coarse-to-fine theories of visual processing where coarse LSF information guides the integration of fine HSF details.

While it is generally assumed that coarse-to-fine integration relies on feedback projections from higher level inferior-temporal or parietal visual cortices, little is known about the nature of such feedback. One possibility is that feedback to early visual cortex is retinotopically organized in a similar fashion as feedforward input. Because LSF and HSF components in natural images are closely correlated in visual space, a retinotopic organization would allow for LSF evoked feedback to be projected onto the same (albeit larger because of larger receptive field sizes in higher level visual cortex) retinotopic locations, which are then reached by the delayed HSF-related feedforward signal. Evidence from animal electrophysiology suggests that LSF feedback and HSF feedforward input to primary visual cortex temporally coincide (Bullier, 2001). Further support for the temporal feasibility of CtF perception comes from a recent study that found response latencies of SF selective clusters in monkey visual cortex (V1 – V4) to reflect CtF organization (Lu et al., 2018). Therefore, retinotopic feedback to early visual cortex presents a feasible mechanism for LSF information to directly influence the processing of the HSFs in close spatial proximity. This is in line with our results for intact versus scrambled images where we observed reduced HSF dominance in response to intact images, but not in response to scrambled images in which LSF and HSF content are not correlated in space. However, purely retinotopic feedback cannot explain the differences in EEG generalization between human and monkey faces. Human faces evoked a stronger LSF dominance early in the trial and a subsequently reduced HSF dominance compared to monkey faces, although in both stimulus categories LSF are equally indexing HSF, and the categories do not vary much in terms of local contrast changes (similar prominence of e.g. eyes and mouth across exemplars). A recent fMRI study (Revina, Petro and Muckli, 2017) showed classification of natural scenes that generalized across spatial frequency, but not

across feedforward and feedback conditions. Because fMRI measures metabolic demands within small voxels, retinotopic feedback should evoke similar patterns to feedforward input from the same stimulus and should thus allow generalization between feedforward and feedback information. The absence of retinotopic generalization in Revina et al. (2017) is inconsistent with simple retinotopic mapping of feedback although this null result has to be considered with caution.

An alternative feedback mechanism might rely on non-retinotopic, higher level information about the stimulus. This notion is supported by a recent study on the macaque face processing hierarchy that shows feedback to carry features of higher level areas (Schwiedrzik and Freiwald, 2017). The authors manipulated stimulus expectation and found that when the expected stimulus sequence was violated, the prediction error measured in low level face selective areas reflected identity specificity and view invariance, properties generally represented in higher level face selective patches. These results indicate that feedback from higher level areas preserves the tuning properties of its origin and makes those properties available as predictions to regions earlier in the visual hierarchy. Such recurrent predictive feedback could explain our results for both intact and scrambled image categories. While the LSF components in scrambled images do not allow for any coherent predictions of HSF content, the intact images do. Further, it has been shown that the processing of human face images strongly relies on LSF information (Goffaux, Gauthier and Rossion, 2003; Goffaux and Rossion, 2006; Goffaux, 2009; Quek et al., 2018). This highly efficient face perception at a glance has been attributed to the strong expertise human observers have developed early in life in processing faces. For monkey faces, the same expertise is usually not achieved (Pascalis et al., 2005). Therefore, the differences in SF dominance patterns between human and monkey trials could indicate the need to integrate more HSF information in the latter case to reduce uncertainty about the image when LSF is not sufficient to form an adequate prediction. In future studies, progressively obscuring content information from experimental stimuli while leaving spatial correlations between LSF and HSF intact could provide a way to more finely investigate to which extent retinotopic proximity of LSF and HSF information on the one hand and higher level content information on the other hand are driving CtF integration. Also note that in our experiment the first 5 cycles (which according to Quek et al. (2018) are sufficient for face detection) were excluded to avoid contrast dissimilarities between SF conditions. LSF information was limited to 5-8 cycles whereas our HSF range (>8 cycles) included what is more commonly referred to as mid-range frequencies. This choice of filter cut-offs might have reduced effect sizes in our study and potentially obscured additional differences in SF dominance over time. Future experiments should test different ranges of SF to allow for a more precise mapping of SF integration.

In summary, our findings indicate that LSF information modulates the processing of HSF details, in line with CtF theories. The reduced HSF dominance in intact human face-, compared to intact monkey face images, along with the evidence about the high-level content of feedback information provided by Revina et al. and Schwiedrzik and Freiwald favors a high-level interpretation of feedback. Although we expect CtF to be the general mode of spatial frequency integration, our experiment only contrasted highly relevant (face-) stimuli with entirely meaningless noise. We therefore cannot conclude from our data that the described effects will indeed generalize to other image categories. Future research including a larger range of stimuli is needed to assess whether reduced HSF dominance in CtF integration could qualify as a general mechanism or whether our results are limited to face processing.

Where exactly the initial LSF processing takes place and hence where the feedback comes from remains unresolved. In the current study, we used EEG-topography based decoding for the assessment of SF dominance, exploiting EEG's high temporal resolution at the cost of spatial specificity. All classifiers were trained on EEG topographies formed by 47 occipital-temporal electrodes, which cover the full range of the visual hierarchy, from early visual cortex to high level category selective regions in the inferior temporal cortex. As the visual signal progresses along the processing hierarchy, different regions become responsible for the amplitude distribution patterns observed on the scalp surface. To reduce potential confounds of stimulus category on SF decoding, we trained all classifiers exclusively on data from scrambled image trials. However, because scrambled images are likely not processed in the same higher-level visual regions, which is reflected in the scalp patterns (see suppl. Figure 1), generalization performance between scrambled and intact image trials might suffer. Supplementary figure 2 illustrates the reduction in generalization performance over trial time as a function of such progressive diversification of cortical sources contributing to the observed scalp pattern. Importantly though, generalization performance remains significantly above chance until about 300ms into testing time. The low spatial resolution of EEG recordings is a limiting factor in our experiment. Conclusions about the potential origin of LSF feedback require an extension of our paradigm to include parcellations of the cortical sources. A promising candidate for the source of feedback in our example are the high-level face-selective visual regions in the occipito-temporal cortex (i.e., fusiform face area or FFA, and occipital face area, or OFA). Both regions have repeatedly been linked to EEG scalp topographies in response to faces at a latency consistent with the N170 ERP component (Yovel et al., 2008; Nguyen and Cunnington, 2014). FFA has also been shown to represent the LSF structure of the face before its finer details (Goffaux et al., 2010). In our own data, the earliest latencies at which we find differences in SF dominance between stimulus categories are compatible with typical N170 latencies, thus potentially reflecting sources in

FFA/OFA. Given the recent advancements in EEG and MEG source reconstruction methods (Henson et al., 2010; Colclough et al., 2015; Farahibozorg, Henson and Hauk, 2018), our paradigm can easily be extended to analyze distinct regions of interest as well as the relationships among them.

In our experiment, SF dominance also varied as a function of the temporal window that the classification model was based on. This indicates that the pattern of information about SF which the classifier extracts changes with time. Although SF classification is robust at a latency of ~90ms to ~560ms post stimulus onset, for models trained at the later portion of this time window the generalizability across time increases. There are several explanations for this. First, visual evoked potentials display high temporal variability across trials, with larger uncertainty later in processing time (Ouyang, Sommer and Zhou, 2016). Due to this temporal smearing, markers of a given neuronal representation appear to be more widely spread across trials, although the representation might be transient in each individual trial. Further, due to EEG's poor spatial resolution, changes in contributions between close-by sources might be missed.

We believe that our approach based on pattern classification and generalization from single processes to integrated processing conditions provides an ideal tool to characterize complex components of human cognition without the need to interfere with the process under study. The principle of recognizing the building blocks of any complex task and subsequently delineating their neural footprints from the integrated whole can be applied to various problems in neuroscience research in which interactions between processes or the integration of components are the target of investigation. This integrative approach may be transferred to other neuroimaging techniques, such as MEG and fMRI.

In conclusion, we found evidence for category-dependent spatial frequency dominance patterns in the processing of broad-band images. Intact images, especially of human faces, evoked patterns consistent with a coarse-to-fine parsing of the visual stimulus. Our results demonstrate the potential of multivariate decoding techniques to separate stimulus features at the level of the neural response, thus facilitating the use of more ecologically valid stimuli.

Acknowledgements

KP, CJ and VG are supported by the Belgian National Fund for Scientific Research (FNRS). StO is supported by the Netherlands Organisation for Scientific Research (NWO), Grant number 453-15-008. We thank Scannexus Maastricht for providing access to the EEG recording equipment and Miriam Heynckes for her help with participant recruitment and recording. We also thank two anonymous reviewers for providing valuable feedback on the initial manuscript.

References

Bar, M. (2003) 'A cortical mechanism for triggering top-down facilitation in visual object recognition', Journal of cognitive neuroscience. MIT Press, 15(4), pp. 600–609.
Bar, M. (2004) 'Visual objects in context', Nature reviews. Neuroscience, 5(8), p. 617.
Bar, M. et al. (2006) 'Top-down facilitation of visual recognition', Proceedings of the National Academy of Sciences of the United States of America. National Acad Sciences, 103(2), pp. 449–454.
Boynton, G. M. et al. (1996) 'Linear systems analysis of functional magnetic resonance imaging in human V1', Journal of Neuroscience, 16(13), pp. 4207–4221.
Bruce, V. and Young, A. (1986) 'Understanding face recognition', British journal of psychology.
Bullier, J. (2001) 'Integrated model of visual processing', Brain Research Reviews. Elsevier, 36(2), pp. 96–107.
Buračas, G. T. and Boynton, G. M. (2002) 'Efficient design of event-related fMRI experiments using M-sequences', Neuroimage, 16(3), pp. 801–813.
Burt, P. J. (1988) 'Smart sensing within a pyramid vision machine', Proceedings of the IEEE. IEEE, 76(8), pp. 1006–1015.
De Cesarei, A. and Codispoti, M. (2013) 'Spatial frequencies and emotional perception', Reviews in the Neurosciences. De Gruyter, 24(1), pp. 89–104.
Chang, C.-C. and Lin, C.-J. (2011) 'LIBSVM: a library for support vector machines', ACM Transactions on Intelligent Systems and Technology (TIST), 2(3), p. 27.
Cichy, R. M., Pantazis, D. and Oliva, A. (2014) 'Resolving human object recognition in space and time', Nature neuroscience. Nature Publishing Group, 17(3), p. 455.
Colclough, G. L. et al. (2015) 'A symmetric multivariate leakage correction for MEG connectomes', NeuroImage. Elsevier, 117, pp. 439–448.
Eimer, M. (2011) 'The face-sensitive N170 component of the event-related brain potential', The Oxford handbook of face perception. Oxford University Press Oxford, UK, 28, pp. 329–344.
Farah, M. J. et al. (1998) 'What is" special" about face perception?', Psychological review. American Psychological Association, 105(3), p. 482.
Farahibozorg, S.-R., Henson, R. N. and Hauk, O. (2018) 'Adaptive cortical parcellations for source reconstructed EEG/MEG connectomes', NeuroImage. Elsevier, 169, pp. 23–45.
Fiorentini, A., Maffei, L. and Sandini, G. (1983) 'The role of high spatial frequencies in face perception', Perception. SAGE Publications Sage UK: London, England, 12(2), pp. 195–201.
Gavrila, D. M. and Philomin, V. (1999) 'Real-time object detection for" smart" vehicles', in Computer Vision, 1999. The Proceedings of the Seventh IEEE

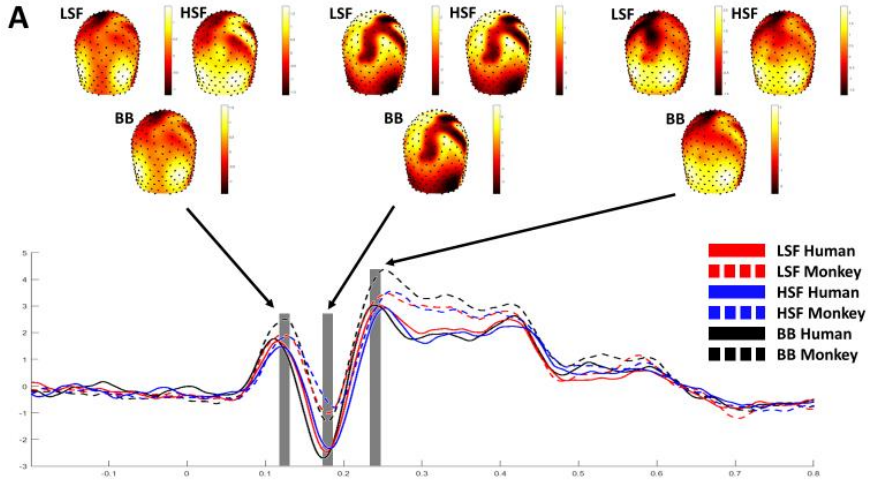
International Conference on. IEEE, pp. 87–93.
Goffaux, V. (2009) 'Spatial interactions in upright and inverted faces: Re-exploration of spatial scale influence', Vision research. Elsevier, 49(7), pp. 774–781.
Goffaux, V. et al. (2010) 'From coarse to fine? Spatial and temporal dynamics of cortical face processing', Cerebral Cortex. Oxford University Press, 21(2), pp. 467–476.
Goffaux, V., Gauthier, I. and Rossion, B. (2003) 'Spatial scale contribution to early visual differences between face and object processing', Cognitive Brain Research. Elsevier, 16(3), pp. 416–424.
Goffaux, V. and Rossion, B. (2006) 'Faces are" spatial"--holistic face perception is supported by low spatial frequencies.', Journal of Experimental Psychology: Human Perception and Performance. American Psychological Association, 32(4), p. 1023.
Halit, H. et al. (2006) 'Is high-spatial frequency information used in the early stages of face detection?', Brain Research. Elsevier, 1117(1), pp. 154–161. doi: 10.1016/j.brainres.2006.07.059.
Haxby, J. V, Hoffman, E. A. and Gobbini, M. I. (2000) 'The distributed human neural system for face perception', Trends in cognitive sciences. Elsevier, 4(6), pp. 223–233.
Hegdé, J. (2008) 'Time course of visual perception: coarse-to-fine processing and beyond', Progress in neurobiology. Elsevier, 84(4), pp. 405–439.
Henson, R. N. et al. (2010) 'A Parametric Empirical Bayesian framework for fMRI-constrained MEG/EEG source reconstruction', Human brain mapping. Wiley Online Library, 31(10), pp. 1512–1531.
Hughes, H. C., Nozawa, G. and Kitterle, F. (1996) 'Global precedence, spatial frequency channels, and the statistics of natural images', Journal of cognitive neuroscience. MIT Press, 8(3), pp. 197–230.
Jones, R. and Keck, M. J. (1978) 'Visual evoked response as a function of grating spatial frequency.', Investigative ophthalmology & visual science. The Association for Research in Vision and Ophthalmology, 17(7), pp. 652–659.
Kauffmann, L. et al. (2015) 'Rapid scene categorization: Role of spatial frequency order, accumulation mode and luminance contrast', Vision research. Elsevier, 107, pp. 49–57.
Kay, K. N. and Yeatman, J. D. (2017) 'Bottom-up and top-down computations in word-and face-selective cortex', eLife. eLife Sciences Publications Limited, 6, p. e22341.
King, J. R. and Dehaene, S. (2014) 'Characterizing the dynamics of mental representations: the temporal generalization method', Trends in cognitive sciences. Elsevier, 18(4), pp. 203–210.
Lu, Y. et al. (2018) 'Revealing Detail along the Visual Hierarchy: Neural Clustering Preserves Acuity from V1 to V4.', Neuron. Elsevier, 98, pp. 1–12.
Marr, D. (1982) 'Vision: A computational investigation into the human

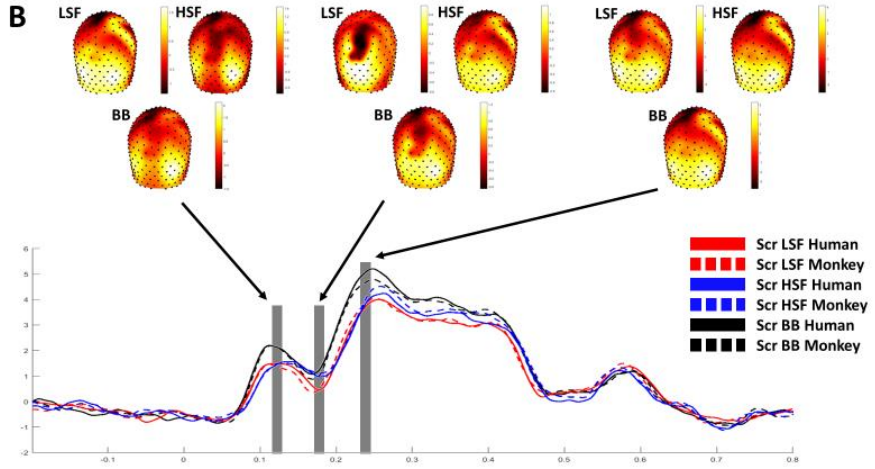
representation and processing of visual information', WH San Francisco: Freeman and Company, 1(2).
Mathworks (2016) 'MATLAB R2016b'. Natick, Massachusetts, United States: The MathWorks, Inc.,.
Mazer, J. A. et al. (2002) 'Spatial frequency and orientation tuning dynamics in area V1', Proceedings of the National Academy of Sciences. National Acad Sciences, 99(3), pp. 1645–1650.
Mihaylova, M., Stomonyakov, V. and Vassilev, A. (1999) 'Peripheral and central delay in processing high spatial frequencies: reaction time and VEP latency studies', Vision research. Elsevier, 39(4), pp. 699–705.
Mu, T. and Li, S. (2013) 'The neural signature of spatial frequency-based information integration in scene perception', Experimental Brain Research. Springer-Verlag, 227(3), pp. 367–377. doi: 10.1007/s00221-013-3517-1.
Musel, B. et al. (2014) 'Coarse-to-fine categorization of visual scenes in scene-selective cortex', Journal of cognitive neuroscience. MIT Press, 26(10), pp. 2287–2297.
Musselwhite, M. J. and Jeffreys, D. A. (1985) 'The influence of spatial frequency on the reaction times and evoked potentials recorded to grating pattern stimuli', Vision research. Elsevier, 25(11), pp. 1545–1555.
Näsänen, R. (1999) 'Spatial frequency bandwidth used in the recognition of facial images', Vision research. Elsevier, 39(23), pp. 3824–3833.
Nguyen, V. T. and Cunnington, R. (2014) 'The superior temporal sulcus and the N170 during face processing: single trial analysis of concurrent EEG–fMRI', Neuroimage. Elsevier, 86, pp. 492–502.
Ojanpää, H. and Näsänen, R. (2003) 'Utilisation of spatial frequency information in face search', Vision research. Elsevier, 43(24), pp. 2505–2515.
Oostenveld, R. et al. (2011) 'FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data', Computational intelligence and neuroscience, 2011, p. 1.
Ouyang, G., Sommer, W. and Zhou, C. (2016) 'Reconstructing ERP amplitude effects after compensating for trial-to-trial latency jitter: A solution based on a novel application of residue iteration decomposition', International Journal of Psychophysiology. Elsevier, 109, pp. 9–20.
Parker, D. M. (1980) 'Simple reaction times to the onset, onset, and contrast reversal of sinusoidal grating stimuli', Attention, Perception, & Psychophysics. Springer, 28(4), pp. 365–368.
Parker, D. M. and Dutch, S. (1987) 'Perceptual latency and spatial frequency', Vision research. Elsevier, 27(8), pp. 1279–1283.
Parker, D. M., Lishman, J. R. and Hughes, J. (1992) 'Temporal integration of spatially filtered visual images', Perception. SAGE Publications Sage UK: London, England, 21(2), pp. 147–160.

Parker, D. M., Lishman, J. R. and Hughes, J. (1997) 'Evidence for the view that temporospatial integration in vision is temporally anisotropic', Perception. SAGE Publications Sage UK: London, England, 26(9), pp. 1169–1180.
Pascalis, O. et al. (2005) 'Plasticity of face processing in infancy', Proceedings of the national academy of sciences. National Acad Sciences, 102(14), pp. 5297–5300.
Peyrin, C. et al. (2006) 'Effect of temporal constraints on hemispheric asymmetries during spatial frequency processing', Brain and cognition. Elsevier, 62(3), pp. 214–220.
Quek, G. L. et al. (2018) 'Ultra-coarse, single-glance human face detection in a dynamic visual stream', NeuroImage. Elsevier.
Revina, Y., Petro, L. S. and Muckli, L. (2017) 'Cortical feedback signals generalise across different spatial frequencies of feedforward inputs', NeuroImage. Elsevier.
Richler, J. J., Cheung, O. S. and Gauthier, I. (2011) 'Holistic processing predicts face recognition', Psychological science. Sage Publications Sage CA: Los Angeles, CA, 22(4), pp. 464–471.
Robson, J. G. (1966) 'Spatial and temporal contrast-sensitivity functions of the visual system', Josa. Optical Society of America, 56(8), pp. 1141–1142.
Schwiedrzik, C. M. and Freiwald, W. A. (2017) 'High-level prediction signals in a low-level area of the macaque face-processing hierarchy', Neuron. Elsevier, 96(1), pp. 89–97.
Schyns, P. G. and Oliva, A. (1994) 'From blobs to boundary edges: Evidence for time- and spatial-scale-dependent scene recognition', Psychological science. SAGE Publications Sage CA: Los Angeles, CA, 5(4), pp. 195–200.
Schyns, P. G. and Oliva, A. (1999) 'Dr. Angry and Mr. Smile: When categorization flexibly modifies the perception of faces in rapid visual presentations', Cognition. Elsevier, 69(3), pp. 243–265.
Sergent, J. (1982) 'About face: left-hemisphere involvement in processing physiognomies.', Journal of Experimental Psychology: Human Perception and Performance. American Psychological Association, 8(1), p. 1.
Sergent, J. (1986) 'Microgenesis of face perception', in Aspects of face processing. Springer, pp. 17–33.
De Valois, R. L. and De Valois, K. K. (1980) 'Spatial vision', Annual review of psychology. Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA, 31(1), pp. 309–341.
Vlamings, P. H. J. M., Goffaux, V. and Kemner, C. (2009) 'Is the early modulation of brain activity by fearful facial expressions primarily mediated by coarse low spatial frequency information?', Journal of vision. The Association for Research in Vision and Ophthalmology, 9(5), p. 12.
Watt, R. J. (1987) 'Scanning from coarse to fine spatial scales in the human visual system after the onset of a stimulus', JOSA A. Optical Society of America, 4(10), pp. 2006–2021.

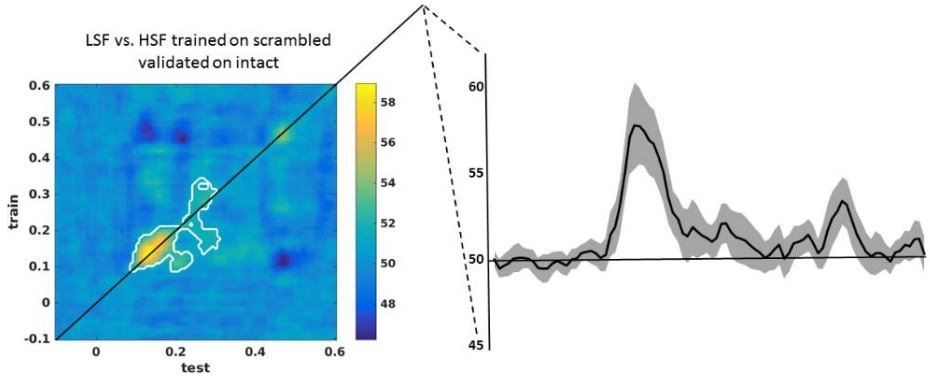
Yovel, G. et al. (2008) 'The face-selective ERP component (N170) is correlated with the face-selective areas in the fusiform gyrus (FFA) and the superior temporal sulcus (fSTS) but not the occipital face area (OFA): a simultaneous fMRI-EEG study', Journal of Vision. The Association for Research in Vision and Ophthalmology, 8(6), p. 401.
Zhang, J. et al. (2014) 'Coarse-to-fine auto-encoder networks (CFAN) for real-time face alignment', in European Conference on Computer Vision. Springer, pp. 1–16.
Zhou, E. et al. (2013) 'Extensive facial landmark localization with coarse-to-fine convolutional network cascade', in Proceedings of the IEEE International Conference on Computer Vision Workshops, pp. 386–391.

Supplementary Figures

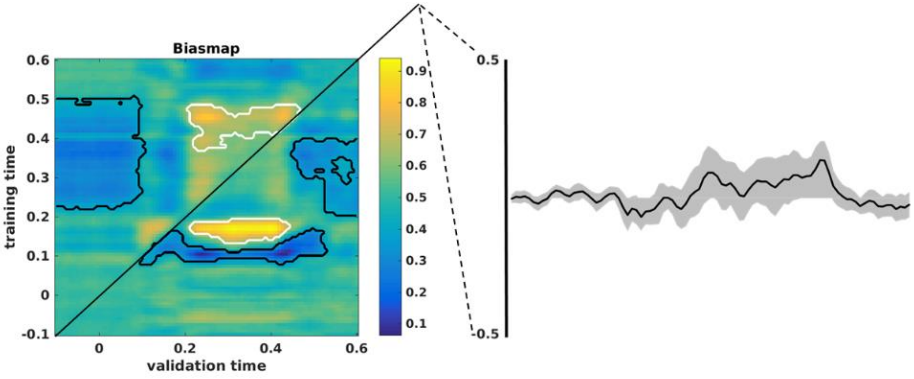




Supplementary figure 1. A) ERP in response to intact images at channel PO8. Topographies are in response to intact human images. B) ERP in response to scrambled images at channel PO8. Topographies are in response to scrambled human images. Not how topographies at the latency of the (face selective) N170 ERP component differ considerably between intact and scrambled conditions. While these differences in topographical distribution of currents measured on the head surface indicates different underlying sources, the representation of spatial frequency still allows for cross-classification at this latency (see suppl. Figure 2).



Supplementary figure 2: Left: Classification performance of classifiers trained on scrambled image LSF vs. HSF trials and evaluated on intact image LSF vs. HSF trials across time. Right: diagonal of classification performance. Classifiers can predict class labels better than chance even when trained on scrambled and validated on intact images, indicating that EEG topographies contain information on the spatial frequency participants saw on a given trial that is at least partially independent of image content. However, classification performance drops after about 200ms. This might be due to intact and scrambled images being processed in different cortical regions.



Supplementary figure 3. Classifier bias in LSF vs. HSF scrambled classification. Left: proportion of HSF predictions across time made by classifiers trained and cross-validated on LSF vs HSF scrambled image trials. Note that the pre-stimulus period is consistently classified as LSF. Right: diagonal of prediction-bias. Note that when classifiers are cross-validated on data from the same time-bins (but independent trials) classifiers show no consistent bias.

Chapter 4

Image coherence modulates feedback to early visual cortex

Kirsten Petras[1,2], Sanne ten Oever[3,4,2], Sarang S. Dalal[5*], Valerie Goffaux[1,2*]
1 UC Louvain, Belgium
2 Maastricht University, The Netherlands
3 Max Planck Institute for Psycholinguistics, The Netherlands
4 Donders Institute for Cognitive Neuroimaging, The Netherlands
5 Aarhus University, Denmark
* Authors contributed equally

Abstract

Visual images contain redundant information across spatial scales. Previous research suggests that the visual system makes use of those redundancies to aid efficient processing. In this framework, a fast initial analysis of low-spatial frequency (LSF) information informs the slower and later processing of high spatial frequencies (HSF). We hypothesize that LSF-guided processing is implemented through LSF-informed feedback from higher-order regions to facilitate the integration of HSF information in early visual cortex. To test our hypothesis, we analysed magnetoencephalography responses to the passive viewing of images of either intact faces or their phase-scrambled version filtered to contain LSF, HSF or LSF + HSF information (broadband). Importantly, only in the intact broadband condition are LSF and HSF correlated (i.e. containing redundant information). We replicated previous findings showing reduced HSF processing for the intact compared to the scrambled condition. This was accompanied by a reduction in the early visual cortex gamma frequency band response, that showed a correlation with response power in the beta range in both fusiform gyrus and frontal cortex. Gamma band oscillations are typically associated with local processing whereas beta band oscillations are interpreted as feedback signals. We therefore interpret our findings as reflecting feedback facilitation of HSF processing in early visual cortex: When LSF is spatially aligned to HSF information, as is the case in natural viewing conditions, LSF-driven feedback from fusiform and frontal regions seems to guide HSF processing in early visual cortex.

Such image-redundancy-related feedback could provide a mechanism for fast and efficient processing of visual objects.

Introduction

Human object recognition is often characterized by a low spatial frequency (LSF) bias at the earliest latencies of visual processing e.g., (Mazer, Vinje, McDermott, Schiller, & Gallant, 2002; Parker & Dutch, 1987). In natural images, low and high spatial frequencies are spatially aligned such that low spatial frequency phase is informative towards the location of high frequency contrast. This property is lost when broadband images are phase-scrambled or when low and high frequencies are presented separately. We have previously shown that in intact, broadband images where coarse (LSF) image contrast is informative towards the high spatial frequency (HSF) content of an image, the human EEG shows a reduction of HSF specific processing markers (Petras, ten Oever, Jacobs, & Goffaux, 2019). We hypothesized that this is due to LSF-driven feedback to early visual cortex (EVC) that serves to restrict HSF processing to sections of visual space that are relevant for recognition. However, both the origin and the information content of such LSF-driven feedback are disputed (see Hegdé, 2008 for review). Within predictive coding frameworks (Friston & Kiebel, 2009; Rao & Ballard, 1999), higher-level areas provide lower level areas with an initial inference about the nature of a stimulus. Lower-level areas match this with incoming sensory evidence and create an error signal in cases of prediction-evidence mismatch. Schwiedrzik and Freiwald (2017) demonstrated that tuning properties of higher-order face-selective patches in macaque cortex can be recovered from prediction errors generated by a hierarchically lower face selective area, indicating that feedback carries information content not originally available to its target area. Here, we hypothesize a similar feedback mechanism in which LSF-evoked feedback provides EVC with information about the relative relevance of spatially distributed HSF information based on an initial assessment of coarse image structure (also see Bar et al., 2006; Bullier, 2001).

If LSF feedback guides HSF processing based on image content information, it likely originates in an area that responds to visual stimuli with receptive fields that are both sufficiently large to cover the global shape of a stimulus and sufficiently abstract to represent complex objects such as inferior temporal cortex or frontal cortex (Bar, 2004; Kietzmann et al., 2019). Inferior temporal cortex is thought to contribute crucially to visual object recognition and shows spatial frequency sensitivity dynamics consistent with a coarse-to-fine integration of spatial scales (Goffaux et al., 2010). Orbitofrontal cortex (OFC) shows a bimodal spatial frequency tuning function (Fintzi & Mahon, 2014) and has also been suggested to be the source of spatial frequency specific feedback to EVC (Bar et al., 2006; Kauffmann, Chauvin, Guyader, & Peyrin, 2015). Here, we use magnetoencephalography (MEG) to separate the responses of these regions and investigate the dynamics of their signaling with EVC. We extract MEG frequency-specific responses, taking advantage

of previous work in humans and monkeys that found feedforward and feedback signals to be represented in different frequency ranges, with feedforward information primarily being signaled in the gamma range and feedback information in the alpha/beta range (Bastos et al., 2015; Fries, 2015; Michalareas et al., 2016; Richter, Thompson, Bosman, & Fries, 2017).

We hypothesize that after a rapid processing by the ventral visual pathway and/or frontal cortex LSF information is fed back to guide primary visual processing to relevant HSF cues. We therefore expect to see the effect of LSF feedback mostly in the neural correlates of HSF processing. To test our hypothesis, we presented participants with images of human faces filtered to contain either LSF only, HSF only or both spatial frequencies (broadband). Images were either intact or phase-scrambled. In intact broadband images, LSF and HSF content is strongly correlated in their spatial distribution, whereas in scrambled versions of the same images, LSF is no longer informative about HSF structure because the image-defining phase structure is lost in the scrambling procedure. Using this design, we can test several predictions regarding the contribution of LSF-driven feedback to visual recognition: First, if the LSF information in an image is informative towards its HSF content (as is the case in broadband natural images), HSF processing is more efficient resulting in reduced HSF specific responses (see also Petras et al., 2019). Second, if LSF guiding alleviates resources from HSF processing, early visual cortex local computation of feedforward signals is reduced, resulting in reduced gamma power over primary visual regions in response to broadband intact compared to broadband scrambled images. And third, if LSF-generated feedback comes from inferior temporal cortex, activity in the alpha/beta bands over inferior temporal cortex is negatively correlated with activity in the gamma band over EVC. Conversely, if frontal cortex generates LSF feedback, a negative correlation between frontal alpha/beta power and EVC gamma power should be found. In a previous study (Petras et al., 2019), we found the predicted reduction of HSF processing markers in response to broadband intact images at a latency of ~300-400ms. If this reduction in HSF dominance stems from a reduction of resources dedicated to HSF, rather than from an increase in resources dedicated to LSF processing, the predicted gamma-band power decrease should be observed at a similar latency.

Results

We investigated the signaling dynamics between EVC and inferior temporal, as well as frontal cortex using MEG. Participants viewed images of intact and phase-scrambled images of faces with similar amplitude and orientation spectra. Images were filtered to contain either low, high, or both (broadband) spatial frequencies. The filter cutoffs were chosen such that LSF and HSF components contributed

equally to the overall contrast in the broadband images (see Methods). Phase-scrambled images were produced by iteratively Fourier-scrambling images and re-pasting the original face to ensure that amplitude and orientation spectra were closely matching those of the original images and not influenced by the uniform gray background (see figure 1 and methods).

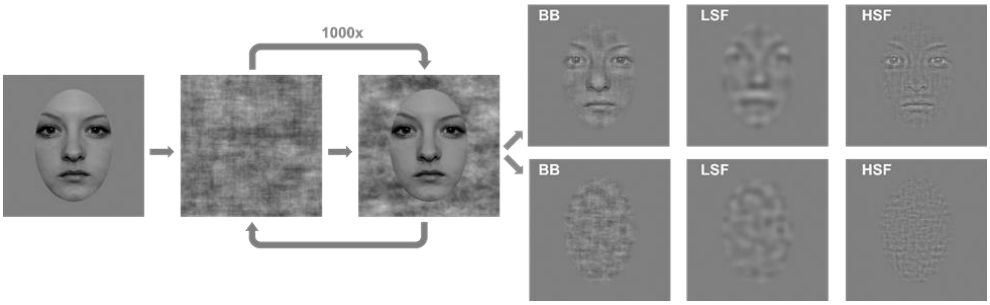


Figure 1: Stimuli were produced from full spectrum, grayscale photographs of human faces by iteratively phase scrambling the image and re-pasting the face to avoid differences in amplitude spectra between intact and scrambled images as well as filter artifacts. LSF and HSF images were produced by multiplying the Fourier spectra with low pass and high pass filters, respectively. broadband images were multiplied with the addition of LSF and HSF filters without contrast normalization.

Sensor space spatial frequency dominance patterns are modulated by image coherence

Tracking the dynamics of spatial frequency integration in naturalistic broadband images is complicated by the need to disentangle the contribution of each frequency band without interfering with the integration itself. We have previously demonstrated that multivariate classifiers trained to distinguish between patterns evoked by either frequency band in isolation can be used to track spatial frequency processing dominance during broadband stimulation (Petras et al., 2019). Here, we replicate these findings in a different group of participants, using MEG instead of EEG.

We trained linear discriminant analysis (LDA) classifiers to perform a time-point by time-point discrimination of LSF and HSF phase-scrambled image trials to avoid any potential influence of image semantic content on classification. The distribution of magnetic field changes measured by a set of 204 planar gradiometers served as features for the classification (see figure 2 for topographies and event related fields). Classification performance was significantly above chance as revealed by a cluster-based permutation test [$p < 0.001$, cluster statistic = 26327 (maximum sum

of cluster t-values)] and first exceeded chance levels around 90ms after stimulus onset (see figure 3). Performance remained above chance until about 450ms after stimulus onset (i.e. 200ms after stimulus offset). As expected, classification performance was highest when training data and cross-validation data were sampled from the same latencies in trial-time as seen along the diagonal in figure 3a. However, there was considerable cross-time generalization (off-diagonal decoding).

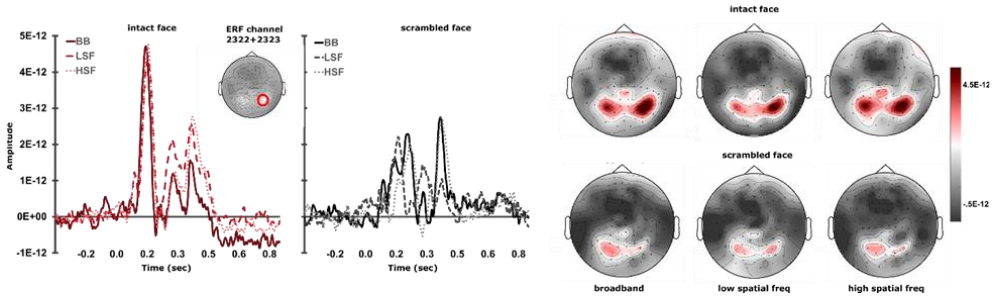


Figure 2: Sensor space event-related fields (ERFs) and topographies for all conditions. Left: ERFs at combined gradiometers 2322 + 2323. Right: Topographies at the latency of the first face selective peak (150ms ±15ms after stimulus onset)

We evaluated how the classifiers pre-trained to differentiate between LSF and HSF scrambled image trials performed on broadband trials. In the latter case the classifiers response reflects which of the two spatial frequency ranges is most dominantly represented in the scalp pattern evoked by the broadband trials. For both broadband intact and scrambled trial data, the classifiers responses significantly deviated from chance towards LSF [Intact face trials: $p_{cluster1} = 0.004$, $p_{cluster2}=0.008$, $p_{cluster3}<0.05$, cluster statistic $_{cluster1} = -6096.12$, cluster statistic $_{cluster2} = -4652.02$, cluster statistic $_{cluster3} = -3704.29$] and [Scrambled trials: $p<0.05$, cluster statistic $= -1834.44$], as well as towards HSF in separate clusters [Intact face trials: $p<0.05$, cluster statistic $= 2351.66$], [Scrambled trials: $p_{cluster1}<0.001$, $p_{cluster2}<0.05$, cluster statistic $_{cluster1} = 11781.78$, cluster statistic $_{cluster2} = 1801.55$]. Importantly, intact and scrambled image trials differ significantly in their spatial frequency dominance with topographies evoked by intact faces showing greater LSF dominance (i.e. lower HSF dominance) than those evoked by scrambled images [$p_{cluster1\&2} <0.001$, $p_{cluster3\text{-}8}<0.005$, cluster statistic $_{cluster1}= -7205.50$, cluster statistic $_{cluster2} = -1972.94$, cluster statistic $_{cluster3} = -401.12$, cluster statistic $_{cluster4} = -356.75$, cluster statistic $_{cluster5} = -241.44$, cluster statistic $_{cluster6} = -175.46$, cluster statistic $_{cluster7} = -101.59$, cluster statistic $_{cluster8} = -90.06$] (see figure 3e).

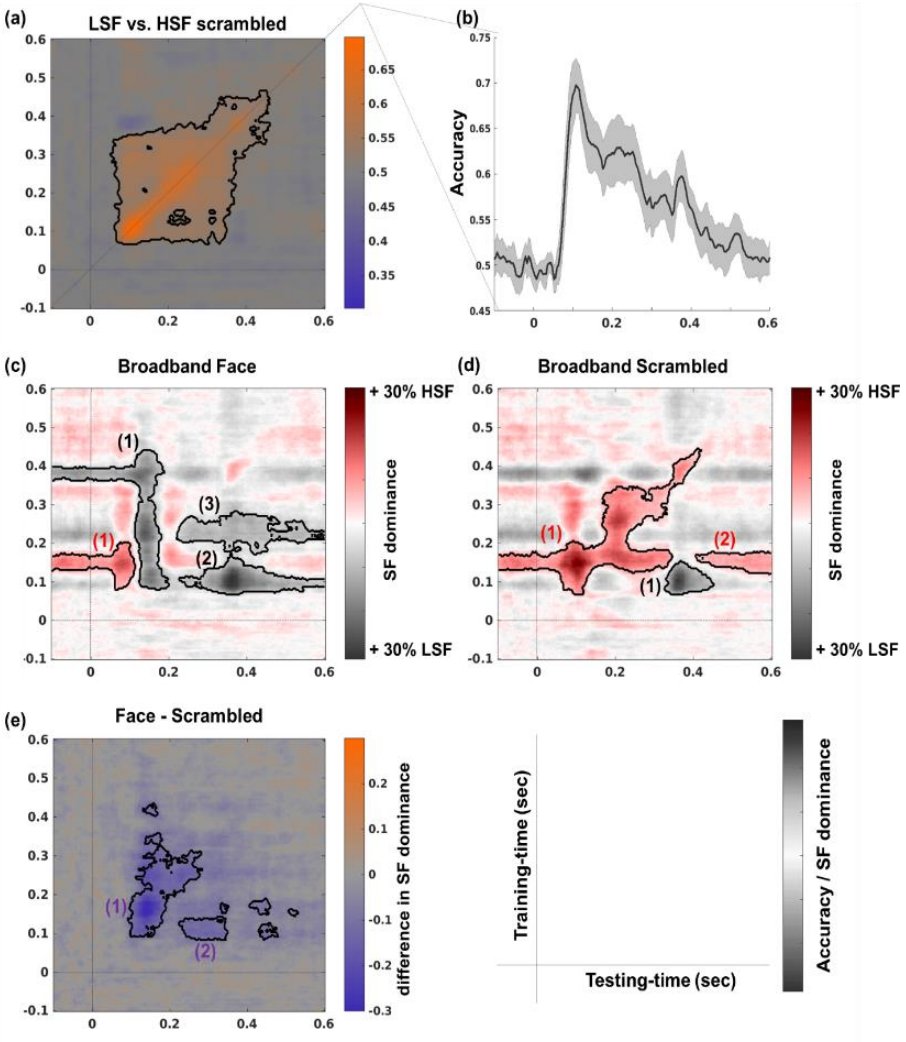


Figure 3: (a) Time by time spatial frequency decoding. (b) Classification performance when training time and testing time are the same (diagonal of (a)). (c) & (d) Generalization performance of classifiers trained in (a) when evaluated on broadband trials. All trials were labelled as HSF for accuracy scoring. Thus, accuracies above .5 indicate the classifier's decision to assign an HSF label, values below .5 indicate an LSF label. (c) Generalization performance for broadband intact face trials (d) Generalization performance for broadband scrambled trials. (e) Difference of classifier generalization between intact face and scrambled trials. Purple colors indicate stronger LSF dominance for intact face trials and stronger HSF dominance for scrambled trials. Cluster numbers correspond to the statistics reported below.

The reduced HSF dominance in response to intact, compared to scrambled images suggests that when LSF aligns in space with HSF, LSF information might guide HSF processing, thus alleviating computational resources.

Functional localizers for individual ROIs

We hypothesized that LSF driven feedback modulates EVC processing when LSF contained information about HSF detail. To constrain our analysis to the relevant cortical regions according to our hypothesis, we performed source reconstructions on our data. We used functional localizer trials to restrict regions of interest (ROIs) to areas that preferentially responded to images of faces. This localization method is robust against small alignment errors, assuming the participants head position was stable within runs (see methods for details).
Prior to each block of the main experiment, participants watched two sequences of face and scrambled image stimuli lasting for 20 seconds each (see methods for details). In order to maximize efficiency and obtain high SNR estimates from limited recording time, we used a periodic visual stimulation technique (Norcia, Appelbaum, Ales, Cottereau, & Rossion, 2015) where images were flickered at a steady rate of 8.5Hz (17Hz with a 50% duty cycle). Such periodic stimulation produces the steady state visual evoked potential (Adrian, 1944; Peli, McCormack, & Sokol, 1988; Strasburger, Scheidler, & Rentschler, 1988), a strong oscillatory potential at the stimulation frequency and its harmonics (see figure 4 for sensor space representation).

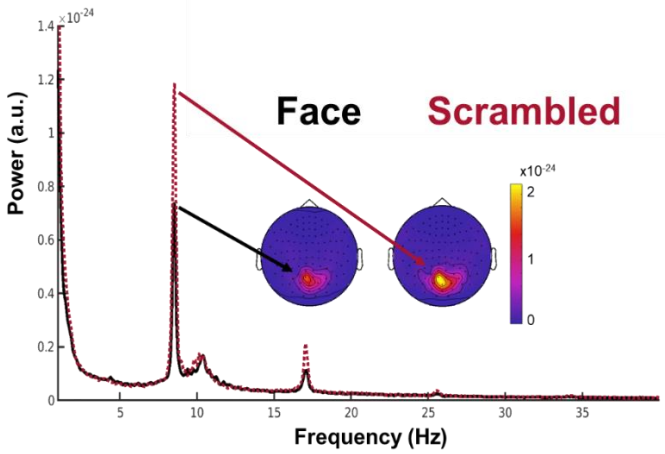


Figure 4: Flickering stimuli at a fixed frequency (here 8.5Hz) evokes a strong response at the stimulation frequency and its harmonics averaged over all occipital channels.

We used dynamic imaging of coherent sources (DICS, Groß et al., 2001)to reconstruct the source level response that was phase-locked to a simultaneously recorded photodiode signal by calculating the coherence coefficient between each grid point in source space and the photodiode signal. We then defined four rough anatomical regions of interest: the calcarine fissure, the lateral occipital cortex, the fusiform gyrus, and the frontal cortex. Within these coarse parcels, we selected the cluster of the 20 source dipoles with the highest summed coherence in response to faces for each participant using k-means clustering (see figure 5 and methods).

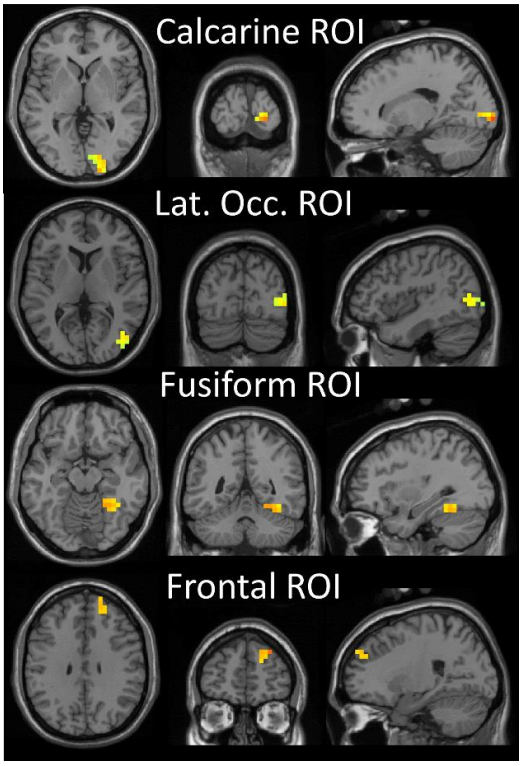


Figure 5: Regions of interest for an example participant. Marked in yellow are the 20 most responsive voxels per ROI. Images are in neurological convention. For all ROIs of all participants see supplementary figure 1.

Source space classification reveals sustained spatial frequency- and image content specific response in early visual cortex

To test our hypothesis that LSF driven feedback reduces the HSF related activation in EVC, we applied the same classification analysis we performed on the sensor space data, on each of the source space ROIs. SF dominance classification cannot be compared to a known ground truth. In contrast to most classification problems, were the classifier output can be evaluated against known class labels, SF

dominance classifiers, that are trained on LSF vs. HSF trials and evaluated on broadband trials, where LSF and HSF contribute equally, are confronted with a classification problem that is ill-posed. To establish that source space classification can generalize across spatial frequency conditions, we therefore first constructed a generalization problem with a known ground truth by classifying between intact and scrambled images within and across SF conditions.

Intact and scrambled image trials can be classified across spatial frequencies

When classifying between intact and scrambled broadband image trials, we found the LDA's performance to be significantly above chance for all investigated ROIs (see table 1).
Classifiers trained on broadband image trial data generalized to both LSF and HSF trial data. Interestingly, generalization performance was significantly higher and widespread for HSF, compared to LSF image trials, indicating a preferential use of HSF information for image type classification.

	BB INTACT VS. SCRAMBLED	**GENERALIZATION TO LSF**	**GENERALIZATION TO HSF**	**DIFFERENCE IN GENERALIZATION PERFORMANCE (LSF-HSF)**
FUSIFORM	$p < .05$, $CS = 9.32 \times 10^4$	$p < .001$, $CS = 3.89 \times 10^4$	$p < .001$, $CS = 1.12 \times 10^5$	(1) $p < .05$, $CS = -1.26 \times 10^4$ (2) $p < .05$, $CS = -4.29 \times 10^3$
LATERAL OCCIPITAL	$p < .05$, $CS = 5.85 \times 10^4$	$p < .001$, $CS = 6.25 \times 10^4$	$p < .001$, $CS = 8.34 \times 10^4$	$p < .05$, $CS = -1.65 \times 10^4$
FRONTAL	(1) $p < .05$, $CS = 6.25 \times 10^3$ (2) $p < .05$, $CS = 4.8 \times 10^3$	(1) $p < .05$, $CS = 2.9 \times 10^3$ (2) $p < .05$, $CS = 2.82 \times 10^3$	$p < .001$, CS = 2.92x104	$p < .05$, $CS = -1.24 \times 10^4$
EVC	$p < .001$, CS = 1.32x105	$p < .001$, $CS = 1.06 \times 10^5$	$p < .001$, $CS = 1.63 \times 10^5$	$p < .001$, $CS = -3.56 \times 10^4$

Table 1: intact vs scrambled image classification results. LDA classifiers were trained on broadband trials to differentiate between intact and scrambled image trials. Classifiers generalized to both LSF and HSF conditions. CS = cluster statistic (maximum sum).

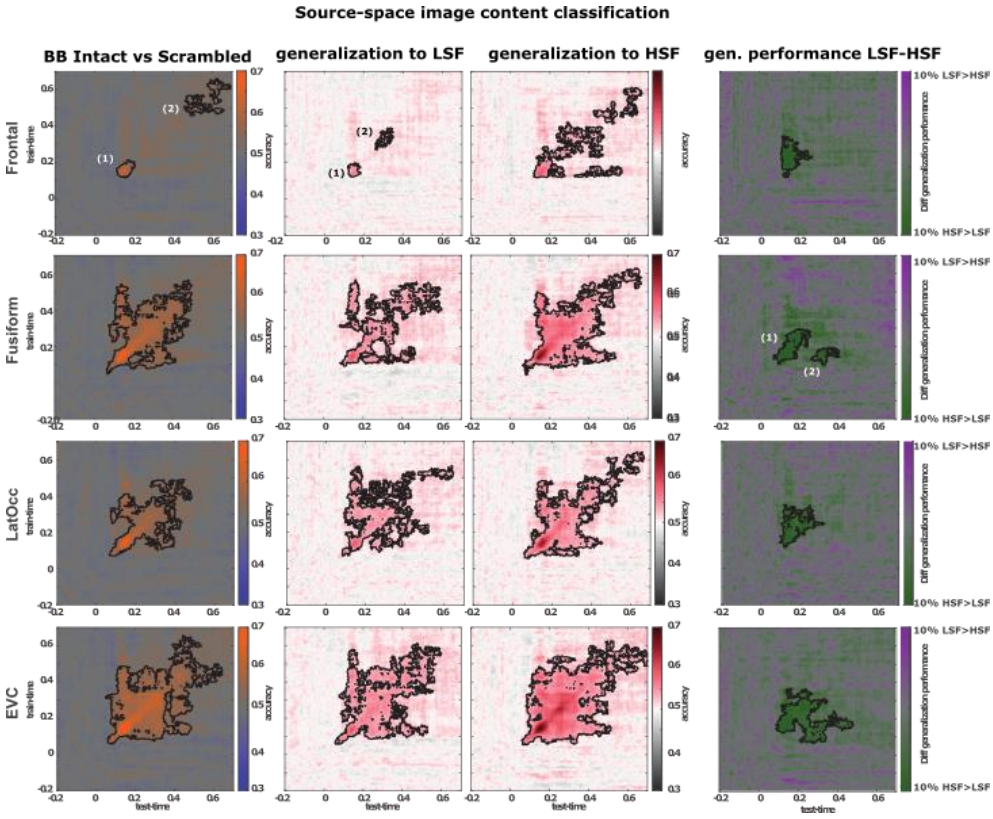


Figure 6: Image content (Face/Scrambled) classification and generalization in source-space. All 4 ROI can differentiate between intact and scrambled broadband image trials. When using the classifiers trained on broadband intact vs. scrambled image trials to categorize intact vs. scrambled LSF and HSF trials, generalization performance is consistently higher for HSF compared to LSF trials. Cluster indices reflect order in ranked cluster statistic (see table 1).

Reduced High spatial frequency dominance for intact images

Classification performance significantly exceeded chance levels in all ROIs when differentiating between low and high spatial frequencies of scrambled stimuli (see table 2).

To reveal which one of the individual spatial frequency bands is most strongly reflected in the activation pattern evoked by broadband trials we applied the classifiers trained on LSF vs HSF scrambled image trials to broadband intact and scrambled image trials. For intact images, both the lateral occipital and the EVC ROI show significant LSF dominance, while the same classifiers applied to scrambled

image trials reveal HSF dominant periods in the Fusiform, Lateral Occipital and the Frontal ROIs, as well as both LSF and HSF dominant periods in the EVC ROI.
We found spatial frequency dominance patterns in early visual cortex to be very similar to those observed in the sensor-space classification indicating a strong contribution of EVC to the pattern observed on the scalp (figure 7). In all investigated ROIs HSF dominance was stronger in response to scrambled, compared to intact images.

	SCRAMBLED LSF VS HSF	**GENERALIZATION TO BROADBAND INTACT**	**GENERALIZATION TO BROADBAND SCRAMBLED**	**DIFFERENCE IN GENERALIZATION PERFORMANCE (IN SCRAMBLED)**
FUSIFORM	$p < .001$, $CS = 4x10^4$	ns ($p > .05$)	$p < .05$, $CS = 1.29x10^4$	$p < .001$, CS = -2.18x
LATERAL OCCIPITAL	$p < .001$, $CS = 7.25x10^4$	$p < .05$, $CS = -7.88x10^3$	$p < .05$, $CS = 1.19x10^4$	$p < .001$, CS =-8.82x
FRONTAL	$p < .05$, $CS = 1.66x10^4$	ns ($p > .05$)	$p < .05$, $CS = 4.92x10^3$	p = .05, CS = -1.42x
EVC	$p < .001$ $CS = 6.79x10^4$	(1) $p < .05$, $CS = -2.85x10^4$; (2) $p < .05$, $CS = 1.52x10^4$; (3) $p < .05$, $CS = -1.24x10^3$	(1) $p < .05$, $CS = 1.78x10^4$; (2) $p < .05$, $CS = 1.56x10^4$; (3) $p < .05$, $CS = -7.22x10^3$; (4) $p < .05$, $CS = 1.78x10^4$	$p < .05$, CS = -2.56x1

Table 2: LSF vs HSF trial classification results. LDA classifiers were trained on scrambled image trials to differentiate between LSF and HSF image trials. CS = cluster statistic (maximum sum). Classifiers generalization bias reflects SF representational dominance with negative values for CS indicating LSF dominance.

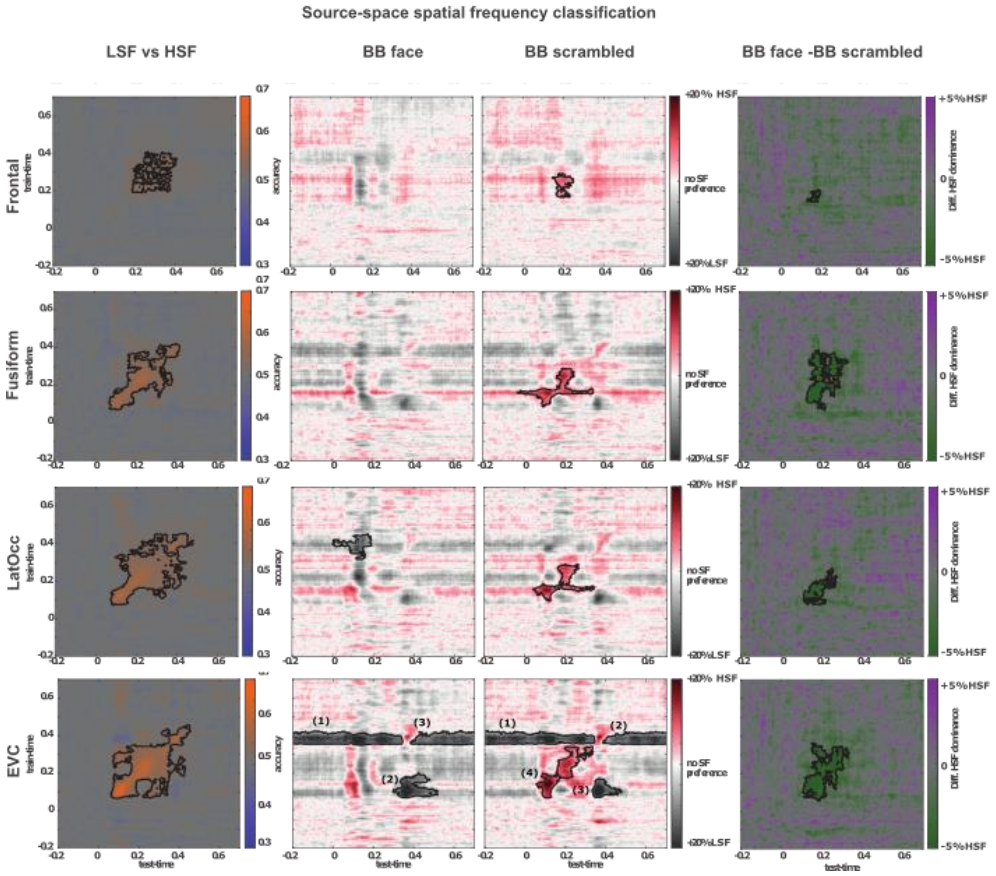


Figure 7: Spatial frequency classification and generalization in source-space. All 4 ROI can differentiate between scrambled LSF and HSF trials, although accuracies and latencies of successful classification differ substantially across them. When applying the classifiers trained on LSF vs HSF scrambled image trials to broadband image trials, the EVC ROI shows a similar SF dominance pattern as observed in sensor-space with stronger HSF dominance in response to scrambled, compared to intact face images. Cluster indices reflect order in ranked cluster statistic (see table 1).

Reduced gamma response in early visual cortex when informative LSF is available

We hypothesized that under conditions that allow for LSF feedback-guiding of HSF processing, i.e. in broadband intact image trials, resources dedicated to HSF processing are reduced. To test this, we determined the power of the neural response in the gamma frequency range as a proxy of local processing for each participant and ROI (see methods for details) and compared gamma responses

across conditions. Figure 8 shows the grand average time-frequency response over all conditions in the Calcarine ROI (see supplementary figure 3 for individual conditions and other ROIs).

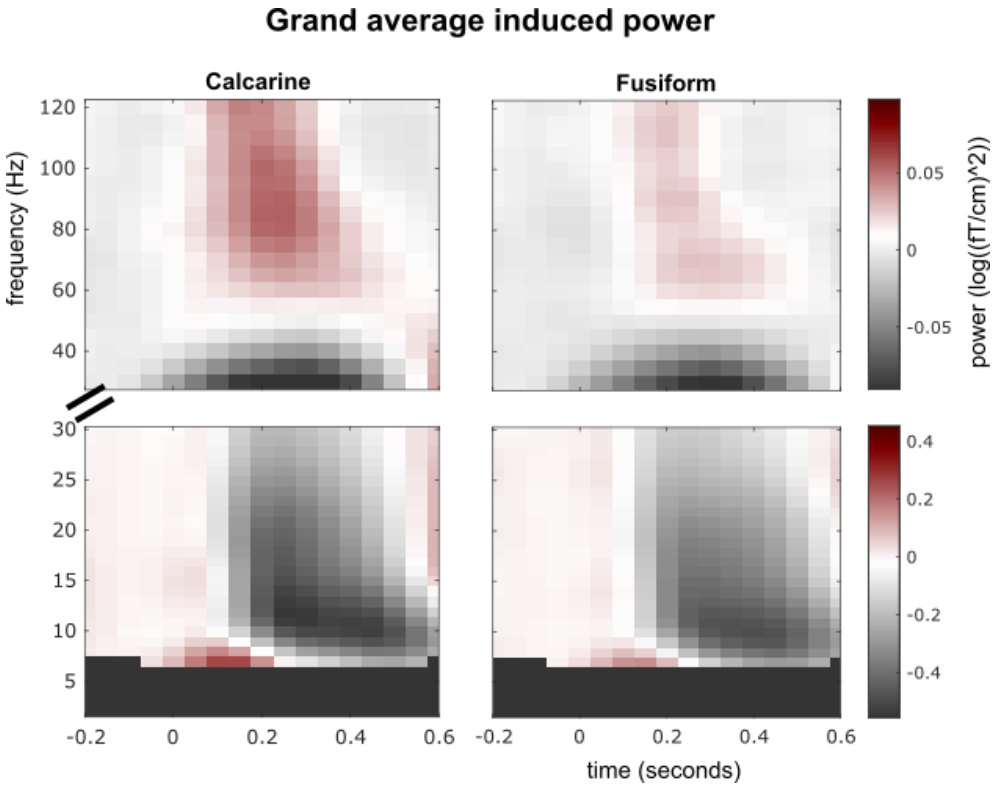


Figure 8: Time-frequency tables of induced power averaged over all conditions in the calcarine (left) and the fusiform (right) ROI. Note that the frequency axis are not continuous between the lower and upper panels.

As shown in figure 9a and b, time frequency analysis revealed lower EVC gamma power for BB, compared to HSF trials [p_{FDR}<.05, latency 250ms] when face images were intact, but not when images were scrambled. This result is striking because broadband images had a higher overall contrast than both LSF and HSF images alone and brain responses to visual stimulation usually increase as a function of contrast (e.g. Gebodh, Vanegas, & Kelly, 2017).

When directly comparing broadband intact and scrambled images, the EVC, but not the Fusiform ROI showed a late decrease in gamma power in response to intact compared to scrambled images [p<.05, FDR corrected for multiple comparisons] (see figure 9 c and d). We interpret this finding as supporting a role of LSF information in the more efficient integration of HSF image detail in EVC. When LSF information contained in the broadband images was informative towards the content of HSF information (i.e. in intact, but not in scrambled images), gamma power decreased. For brevity only the main results are plotted here. See supplementary figure 2 for an overview of all conditions.

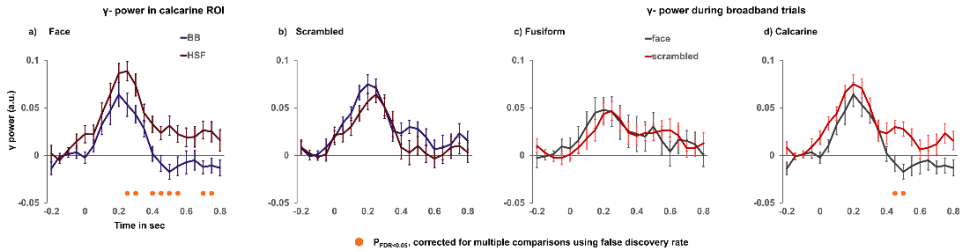


Figure 9: Reduced gamma power in the presence of meaningful LSF information. a) γ-power in EVC is significantly lower in response to broadband, compared to HSF stimuli only when images are intact. b) While the Fusiform ROI shows no significant differences in γ-power between intact and scrambled image trials, which is surprising given the face selectivity of the region, the Calcarine ROI shows a reduction in γ-power for intact, compared to scrambled image trials. Note that the data in 7c is the replotted direct comparison between the broadband conditions in panels a) and b). Error bars indicate standard error of the mean.

Reduced gamma response in early visual cortex correlates with increased beta power in Fusiform and Frontal ROIs for intact images

To investigate the hypothesis that feedback from high-level regions is driving the gamma power decrease in EVC in response to intact compared to scrambled images, we correlated the power of neural responses in alpha and beta bands of high-level regions as a proxy of feedback strength with the gamma response power of EVC at time lags ranging from -100ms (α/β leading) to +100ms (γ leading). We found no difference in alpha/gamma power correlations between intact and scrambled image trials (figure 10). In contrast, beta power showed a stronger negative correlation with EVC gamma power for intact, compared to scrambled image trials in both the Fusiform and Frontal ROIs [figure 10B; Fusiform: $p_{corr\ intact - corr\ scrambled}$ <.05, cluster statistic = -19.82; Frontal: $p_{corr\ intact - corr\ scrambled}$ <.05, cluster statistic = -20.63]. These results indicate that feedback from both Fusiform and the Frontal ROI might be driving the decrease of EVC response in the gamma range.

Although we expected negative correlations to be stronger when examined at lags consistent with feedback to EVC (thus α/β leading), we found no significant differences between lags. However, the latencies (~350ms±50ms) at which the negative correlation between Fusiform and Frontal beta and EVC gamma is stronger for intact compared to scrambled images matches those of the EVC gamma response decrease in response to intact images (compare figure 9d). Interestingly, the Frontal ROI shows the strongest correlation difference between intact and scrambled conditions about 50ms later than the Fusiform ROI. However, note that time bins used in this analysis were wide (50ms) and don't allow for an exact localization of effects in time. Further, the cluster-based permutation testing used for statistical analysis of the correlation differences does not permit the interpretation of cluster latencies as effect latencies (Maris & Oostenveld, 2007). We therefore refrain from drawing strong conclusions from latency effects.

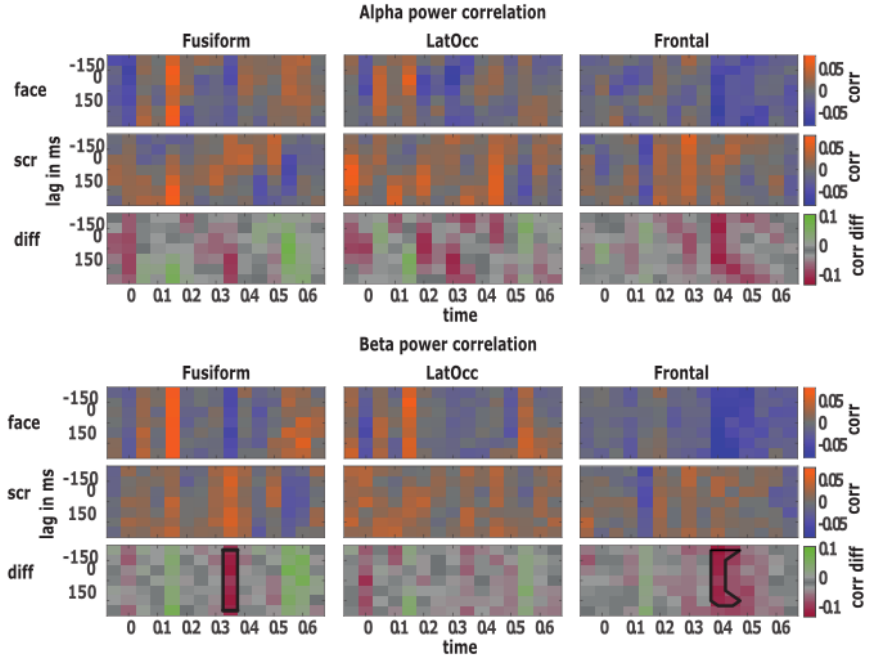


Figure 10: Lagged correlations between high-level ROIs alpha and beta power and EVC gamma power. While there are no significant differences in alpha to gamma power correlations both Fusiform and frontal beta power shows a stronger negative correlation with EVC gamma power when images are intact rather than scrambled. Negative lag indicates that earlier time bins of the higher-level regions were correlated with later time bins of EVC. Positive lags indicate that later time bins of higher-level regions were correlated with earlier time bins of EVC. Lag 0 corresponds to the same time bin in higher-level cortex and EVC.

To investigate the role of LSF feedback in the observed EVC gamma suppression, we compared cross ROI, cross frequency power correlations between broadband and HSF trials. We find a stronger positive correlation for broadband intact images in both alpha and beta with gamma power only for the Fusiform ROI [figure 11; Fusiform alpha: $p_{corr\ broadband - corr\ HSF} < .05$, cluster statistic = 22.23; Fusiform beta: $p_{corr\ broadband - corr\ HSF} < .05$, cluster statistic = 20.38]. The latency of the maximum difference between spatial frequency conditions is consistent with the initial processing of the face. When images are scrambled, no difference in correlation between spatial frequency conditions is found. We interpret this finding as evidence that LSF is a necessary driver for the reduction in EVC gamma response seen with intact images.

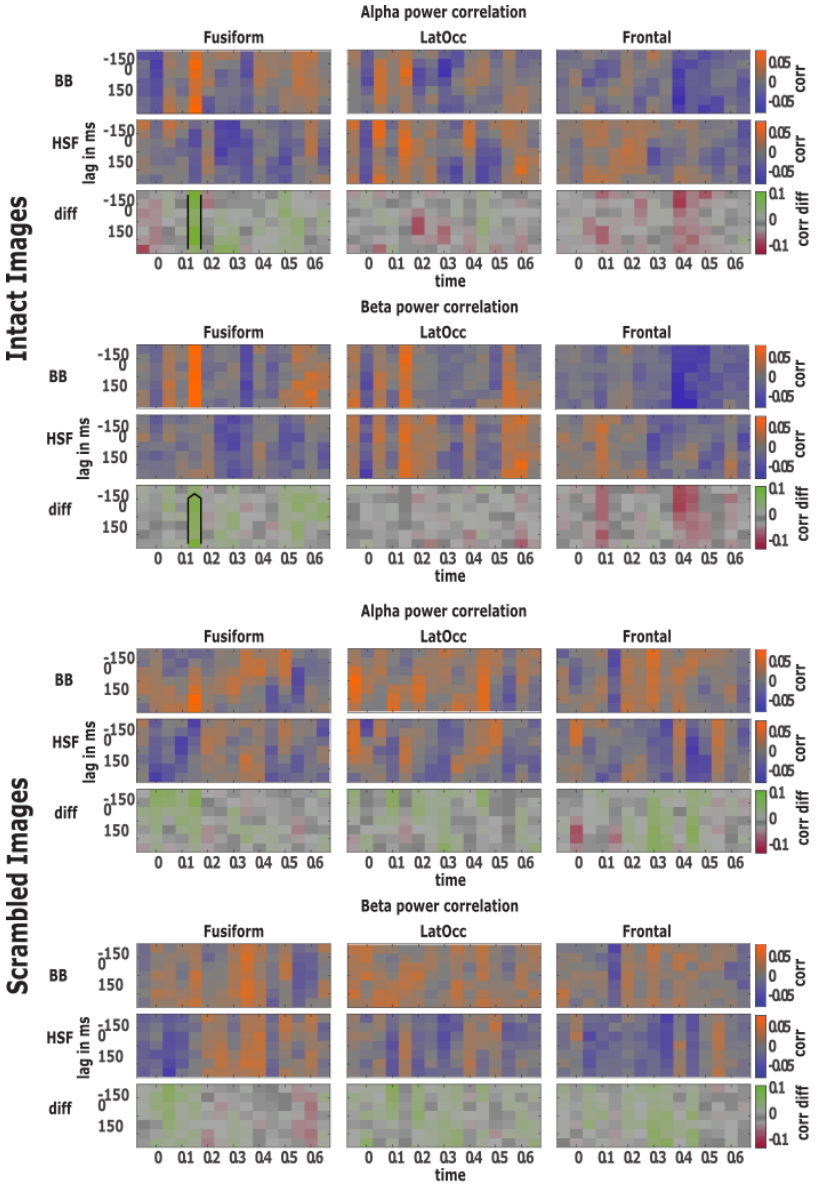


Figure 11: Correlations between high-level ROIs alpha and beta power and EVC gamma power. Fusiform alpha and beta power correlates significantly more with EVC gamma power when images are broadband. Negative lag indicates that earlier time bins of the higher-level regions were correlated with later time bins of EVC. Positive lags indicate that later time bins of higher-level regions were correlated with earlier time bins of EVC. Lag 0 corresponds to the same time bin in higher-level cortex and EVC.

Discussion

Visual objects are characterized by a strong phase congruence across spatial frequencies (e.g. MaBouDi, Shimazaki, Amari, & Soltanian-Zadeh, 2016) such that the phase of low spatial frequency is a good predictor of the location of relevant high spatial frequencies. Fourier phase-scrambling disrupts that relationship between spatial frequency ranges while leaving other first-order statistical regularities such as mean luminance, contrast and orientation spectrum unaffected (see for example Honey, Kirchner, & VanRullen, 2008). Thus, any effect of phase-scrambling on the efficiency of HSF integration necessitates a sufficiently high-level of abstraction from those primary stimulus properties. Traditional views consider V1 cells as mainly driven by contrast, orientation and spatial frequency of the FF input falling within their receptive field. In this framework, phase-scrambling of images should have limited effects within V1 but substantially affect processing in higher-order visual areas that are sensitive to more complex stimulus properties such as shape or semantic content. Conversely, if feedback shapes V1 function, tuning properties of higher-level areas could propagate back and be detected in V1. Note however, that the disruption of continuous local edges introduced via phase-scrambling might also lead to differences due to lateral inhibition processes that do not require feedback (see e.g. Gregor, Szlam, & LeCun, 2011).

Here, we investigated how the spatial correlations between low and high spatial frequencies in intact and scrambled naturalistic images matched for basic V1 tuning properties influence the processing of high spatial frequency detail. We have demonstrated that image coherence influences the integration of low and high spatial frequency information. Consistent with our previous EEG results, we show that sensor space spatial frequency classification of broadband image trials shows a stronger HSF bias when image phase is scrambled to render images of human faces unrecognizable, than when phase remains intact. A similar pattern can be observed in source-reconstructed EVC. Further, in EVC, the power of neural responses in the gamma range, which presumably reflects local computations, is lower in response to broadband intact than to broadband scrambled images. Although in our design broadband images contain higher contrast than either HSF or LSF images alone, EVC gamma power is lower in response to intact broadband images compared to HSF images. This finding is in stark contrast to a purely feedforward interpretation of EVC function, where response properties only depend on the local tuning to basic stimulus properties (Marr, 1982). However, gamma-power changes between conditions with different contrast need to be considered with caution as not only gamma power but also gamma frequency increases with stimulus contrast (Hadjipapas, Lowet, Roberts, Peter, & De Weerd, 2015). We determined individual subjects' gamma frequencies based on the response in each ROI, averaged over all conditions. As the trial numbers for all

conditions were counterbalanced in the experiment, the two lower contrast conditions (LSF and HSF) contributed more trials to the average than the higher contrast broadband condition. It is therefore possible, that the frequency of the broadband gamma response was not optimally represented in the individual gamma frequency ranges we used in our analysis.

Further, beta band activity in Fusiform and Frontal ROIs, taken here as proxy of feedback signaling, shows a stronger negative correlation with EVC gamma power during the viewing of intact compared to scrambled images. Taken together, our results indicate that the availability of informative low spatial frequency determines the processing of high spatial frequency detail. Several mechanisms could account for those outcomes.

The efficient coding theory (H. Barlow, 2001; H. B. Barlow, 1961; Simoncelli, 2003) conjectures that the visual system aims to reduce redundancy by taking statistical regularities of the visual environment into account. Any consistent spatial relationship across spatial frequency ranges, such as LSF indexing of HSF contrast represents redundant information. An efficient way to integrate spatial frequencies would thus be to focus processing resources on locations that are likely relevant. When e.g. identifying faces previous research has shown that the most relevant features are the eye- and mouth regions (Janik et al. 1978; Miellet, Caldara, and Schyns 2011; but see also Blais et al. 2008). LSF contrast is a reliable indicator towards the location of those regions, so favoring HSF processing within high LSF contrast regions would allow for effective and resource-efficient processing of face identity. Such a process could in principle occur at any level of the visual hierarchy where receptive fields from various spatial frequency ranges are converging to allow for input from different spatial scales to be matched. Though, to restrict HSF processing to relevant image regions, at least the coarse retinotopic organization of the stimulus needs to be preserved in the feedback signal. Previous studies that explicitly investigated the spatial distribution of feedback effects to primary visual cortex however, found evidence to the contrary: Williams et al. (2008) found feedback to foveal retinotopic areas of V1 to contain information about objects presented in the periphery. De Wit et al. (2012) found that V1 fMRI BOLD responses to bi-stable stimuli that could either be perceived as 4 individually moving lines or be perceptually organized into a single Gestalt differed dependent on the participants' percept with lower BOLD responses to the global Gestalt percept than to the local individual lines percept. Interestingly, the entire V1 showed this reduction in BOLD amplitude, not only the retinotopic regions corresponding to the perceptual change. Even more strikingly, Hsieh et al. (2010) found evidence that early retinotopic cortex response patterns to two-tone images were more similar to the patterns evoked by the same images' grayscale photograph version if the participant knew the original image than it was to the identical but unrecognized two-tone image. Combined, these findings contradict a role for purely spatially

defined mechanisms such as the here proposed LSF-driven feedback. Due to the low temporal resolution of all three MRI studies, it is difficult to dissociate possible earlier retinotopic effects from possible later more global effects. In (Kok, Jehee, & De Lange, 2012), the authors argue that prior expectation of a relevant stimulus feature sharpens that features' neural representation under noisy conditions. In line with this logic, suppressive feedback to regions of EVC that do not match the retinotopic location of the presented stimulus could reduce stimulus ambiguity and would be measured as an overall decrease in response along with an enhancement of the stimulus-specific representation.

Using the superior temporal resolution of MEG recordings, we find the effects of image phase-scrambling to be relatively late in the brains' response (EVC gamma power reduction for intact vs scrambled around 400ms), although an exact temporal localization would require different tradeoffs between time and frequency resolution than the ones we chose here. The effect of LSF presence as seen in intact broadband compared to intact HSF images occurs earlier (EVC gamma power reduction around 200ms, see figure 6). However, two major limitations of the current design need to be considered here: first, the increase in temporal resolution afforded by MEG in comparison with fMRI recordings comes at the cost of reduced spatial specificity and feedback signals found with MEG cannot currently be linked to a precise retinotopic target within EVC. To non-invasively dissociate between early, potentially retinotopic effects and later, potentially more global effects, the high spatial resolution fMRI should be combined with a temporally sensitive stimulus presentation paradigm such as temporal masking (e.g. Goffaux et al., 2010). Alternatively, temporally sensitive recordings such as MEG or EEG could be combined with spatially distinct stimuli presentation methods. A second concern is that the gamma-frequency-range we chose for our comparisons might not be optimal. To ensure that LSF and HSF images were fully represented in our broadband stimuli, we constructed broadband images by adding the LSF and HSF without normalization. This procedure leads to broadband images having higher RMS contrast than both LSF and HSF alone. Hadjipapas et al. (2015) cautioned that the frequency of gamma responses increases with stimulus contrast. It is therefore possible that our approach to determine individual gamma frequency based on each participant's average response over all conditions, deviated from the optimal gamma frequency for each individual condition. Nevertheless, we find it unlikely a potential increase in gamma frequency in the broadband condition would have been sufficiently severe to remove the response of interest out of our chosen individual gamma range (individual gamma peak ± 10Hz). In the above-mentioned study, the average frequency-increase the authors found, even after correcting for the dominating power increase, was 8HZ for a contrast difference of 60%. The RMS contrasts of our stimuli are ~0.02 for narrowband and ~0.028 for broadband images.

In conclusion, we find evidence for the LSF guiding of HSF image information that is expressed in both, spatial frequency dominance patterns and EVC gamma power. We further find evidence that such LSF guiding might be implemented through feedback connections from higher level visually responsive areas to EVC.

Methods

Participants:

21 participants (9 female, mean age 28.3 SD 5.13) took part in the study in exchange for monetary compensation. Optical corrections in the form of lenses glued to the top of the MEG helmet were used if needed. All participants provided written informed consent and were fully informed about the purpose of the experiment after participation. The study was approved by the local ethical committee of Aarhus. One participant was excluded due to excessive noise in the MEG data and one participant was excluded for failing to reach sufficient task performance.

Stimuli

Stimuli were 47 images each of human faces, houses and their phase-scrambled versions. Houses were selected from the Pasadena houses database (Perona, 2000). We chose the 47 images that best matched the orientation spectrum of our face stimuli. Faces and houses were cropped to the same oval occlusion and pasted onto a uniform gray background. All image luminance histograms were matched to the same reference histogram (the average of all images) using the shine toolbox (Willenbockel et al., 2010) in Matlab. To avoid filter artifacts due to the uniform gray background, the background was replaced with a Fourier phase-scrambled version of the image by first scrambling the whole image and then re-pasting the non-scrambled aperture. Scrambled images were produced in the same way, except that the last step of re-pasting the intact aperture was skipped. To ensure that the scrambled background portrayed the low-level properties of the aperture rather than a mixture of the aperture and the original gray background the process of scrambling and re-pasting the aperture was repeated 1000 times per image thus gradually reducing the influence of the original background on the spatial frequency distribution of the scrambled background. To restrict image content to LSF or HSF, we used 7^{th} order Butterworth filters with filter cutoffs determined such that the area under the curve of the amplitude spectrum would be as similar as possible below and above the cutoff. To avoid very low frequencies accounting for most of the contrast in the LSF images, we excluded the first 5 cycles per image prior to determining the cutoff. This resulted in LSF passbands of > 5 and < 7cpi (>.5 and <.7 cpd horizontal) and HSF cutoffs of >7cpi (>.7 cpd horizontal). Finally, we masked out the phase-scrambled background with a cosine tapered mid-gray background

(see figure 1). Broadband images were constructed as the non-normalized addition of LSF and HSF images. Consequently, broadband images had higher contrast than either LSF or HSF images and importantly, because LSF and HSF images contained the same overall contrast (RMS contrast ~0.02), they contributed equally to the contrast of broadband images (RMS contrast ~0.028).

Experimental procedure

Images were projected centrally onto a screen at 84cm viewing distance from the participant by a ProPixx projector (VPixx Technologies, Canada) in a dimly lit MEG recording chamber. The oval aperture of each image covered 22cm in height and 15cm in width resulting in a visual angle of 15° vertically and 10° horizontally.
Each participant completed 18 localizer trials and 9 experimental blocks. In localizer trials a sequence of images of one category (faces, houses or scrambled) was shown to the participant at a temporal frequency of 8.5 Hz (alternating image/blank screen with a 50% duty cycle). Images were band pass filtered in 18 logarithmically spaced steps covering the full range from LSF low cutoff to the HSF high cutoff. Each localizer trial consisted of a sequence of images over a time period of 18 seconds with increasing or decreasing spatial frequency (preceded by a 1sec ramp-up and followed by a 1sec ramp-down period). Participants were instructed to passively view the sequences.
Experimental blocks consisted of 100 trials each including 10% target trials to assure participants' attention. Each trial started with a fixation dot on a uniformly gray screen that lasted for a jittered interval of 1.5-2.5 seconds. After that, an image of a face, a house or a phase-scrambled version of either of these was presented for 250ms. Each image was filtered to contain either high, low or broadband spatial frequencies using the above described Butterworth filters. Image presentation was followed by a blank period of 1.5 seconds. Participants were instructed to detect interspersed images showing a dog (could be in any of the three spatial frequency conditions) and report detection via a button press with their right hand. The mean luminance of the screen was kept at 350 cd/m^2 throughout the experiment and participants had short breaks between every block as well as a longer break after every three blocks.

Electroretinogram

We recorded the electroretinogram (ERG) from both eyes using a corneal silver/nylon electrode (DTL Plus Electrode, Diagnosys, USA). The electrode was placed on the participant's cornea such that it rested on the lower eyelid. Presumably due to the low contrast at high overall luminance in our stimulation paradigm, we found no discernable ERG response. Data from this recording is therefore not shown here.

MEG acquisition and pre-processing

MEG data were acquired in a magnetically and sound shielded room from 204 planar gradiometers and 102 magnetometers (Elekta Neuromag TRIUX) at Aarhus University at a sampling rate of 2000Hz. Eye movements were tracked via the bilateral horizontal and the right vertical EOG for later use in MEG artifact correction.
Prior to the MEG recording the position of 4 HID coils and the participant's fiducial points as well as their individual head shape were digitized using the Polhemus Fastrak® digitizer (Colchester, VT). Participants were sitting in upright position throughout the recording.
Data were offline notch-filtered to remove line noise and down-sampled to 500Hz using MNE python (Gramfort et al., 2013, 2014). Stimulus triggers were corrected to the signal of a simultaneously recorded light sensitive diode to account for delays between stimulus trigger and the first stimulus frame. Data were segmented to trials reaching from -1500 to 1500 ms around stimulus onset and visually inspected for squid jumps and obvious ocular or muscle artifacts. Bad channels, artifactual trials as well as target trials were excluded from subsequent analyses.

MRI acquisition and pre-processing

T1 weighted MRIs with an isotropic voxel size of 1 mm^3 and a reconstructed matrix size of 256 × 256 of 15 participants were acquired via a 3 T Siemens Skyra scanner at CFIN, Aarhus University, Denmark. The remaining 4 participants provided an equivalent recent T1 weighted MRI recorded for a different study. MRIs were converted to nifty and aligned to AC/PC space using SPM 12 (Penny, Friston, Ashburner, Kiebel, & Nichols, 2011).
Sensor space analysis
Preprocessed trials were low-pass filtered by a 6^{th} order Butterworth zero phase filter with a cutoff frequency of 80Hz, down-sampled to 250 Hz and reduced to 700ms segments ranging from -100ms to 600ms around stimulus onset. Missing channels were interpolated using spherical spline interpolation of neighboring channels and planar gradients over both orientations were combined into a single magnitude measure. Data were baseline corrected to the 100ms before stimulus onset and temporally smoothed to the moving average over three timepoints (12ms). Regularized Linear Discriminant Analysis (LDA) using MVPA light (Blankertz, Lemm, Treder, Haufe, & Müller, 2011) integrated with fieldtrip (Oostenveld, Fries, Maris, & Schoffelen, 2011) was performed on a timepoint x timepoint basis (King & Dehaene, 2014) to classify between LSF and HSF scrambled image trials. We used 5 fold cross-validation and shrinkage regularization as $(1-\lambda)*Sw+\lambda*\nu*I$ where λ is determined by the Ledoit-Wolf estimate (Ledoit & Wolf, 2004), Sw is the scatter matrix, I the identity matrix, P is the number of features and ν is the trace

(i.e. the sum of eigenvalues) of Sw divided by P. The trained classifiers were then evaluated on their generalization performance to broadband intact and scrambled image trials. Since the classifiers have only learned to categorize trial data as belonging to either LSF or HSF trials they will necessarily be incorrect on all broadband trials because broadband images have been specifically designed to contain equal contrast in both low and high spatial frequencies. We therefore interpret the classifier's decision to assign either an LSF or an HSF label to a broadband datapoint as reflecting spatial frequency dominance, thus the MEG topography at the datapoint in question is more similar to one evoked by LSF or HSF stimulation, respectively. Statistical significance of classification results was determined using cluster based permutation testing with alpha set to 0.01 and performing 1,000 randomizations (for details see Maris & Oostenveld, 2007).

MEG source reconstruction

3 head position indicator (HPI) coils were used to determine the participants' head position relative to the MEG sensors during recording. MEG data were aligned to MRI space in a three-step procedure: First the fiducial points provided by the Polhemus Fastrak® digitizer were used to calculate an initial transformation matrix and second the digitized head-shape was manually aligned to a surface reconstruction of the individual's MRI. Finally, an iterative closest point matching procedure was performed. All transformation matrices were inverted, combined into a single transformation-matrix, and applied to the MEG gradiometer descriptions separately for each run. A three compartment boundary element headmodel (brain, skull, scalp) was created using the OpenMEEG implementation in fieldtrip (Gramfort, Papadopoulo, Olivi, & Clerc, 2010; Kybic et al., 2005; Oostenveld et al., 2011). For each participant and run leadfields for ~42000 positions based on a sourcemodel template warped to the individual participants' brain were created using standard pipelines in fieldtrip.
Data from the experimental trials were reconstructed using a linear constrained minimum variance (LCMV) beamformer. We created common spatial filters using data of all conditions.

ROI analysis

Although MEG source reconstruction has repeatedly proven to produce reliable results at spatial resolutions comparable to the one used here (Belardinelli, Ortiz, Barnes, Noppeney, & Preissl, 2012; S. Dalal et al., 2013; Hillebrand & Barnes, 2011), mistakes in the alignment of MEG sensors to the participants' MRI can cause considerable spatial inaccuracies (S. S. Dalal, Rampp, Willomitzer, & Ettl, 2014; Hillebrand & Barnes, 2003, 2011). Thus, while atlas based, anatomically defined regions of interest (ROIs) are convenient, they might not reflect the most

responsive regions in a given subject, due to (1) potential misalignment of the MEG sensors with the MRI and (2) inter-individual differences in functional maps. Here, we produced individual functionally defined regions of interest for each participant using independent localizer data. Each of the localizer trials consisted of images of either faces or their phase-scrambled versions, flickered at an effective frequency of 8.5Hz (17Hz with a 50% duty cycle) and increasing or decreasing in spatial frequency in 18 steps over a time period of 18 seconds (preceded by a 1sec ramp-up and followed by a 1sec ramp-down period). Since the brains' response to periodically changing stimulation is expected to show the same periodicity as the stimulus (Adrian, 1944; see Norcia et al., 2015 for review), we used a single frequency dynamic imaging of coherent sources beamformer (Groß et al., 2001) to localize the grid points with the highest selective response to faces. An underlying assumption of the beamformer is that sources of interest are temporally independent. Because the flicker stimulation results in very a high SNR signal that leads to highly correlated near zero-lag bilateral activation in visual cortices, this assumption is violated, potentially causing inaccurate localization. To nonetheless yield a reliable result from the beamformer, we scanned the brain volume using pairs of left/right hemisphere symmetric dipoles as the source model, a solution suggested by Popov et al. (2018). We then restricted analysis to 4 anatomically defined rough ROIs (calcarine fissure, fusiform gyrus, lateral occipital cortex and frontal cortex) and calculated the coherence between sources during face-image trials and the signal of a light sensitive photodiode during those same trials, normalized by the coherence between sources and diode during scrambled image trials. We used k-means clustering with $k = \frac{n\ voxels}{15}$ on the Euclidean distance between the 3 position coordinates and the coherence values (z-scored to ensure similar scaling of all dimensions) to parcellate sources into 4 dimensional clusters (3 position + 1 coherence dimension). To ensure matching sizes across ROIs and participants, we selected the 20 voxels with the highest coherence score from the cluster with the maximum summed coherence for further analysis. Single trial ROI virtual sensor responses were obtained by multiplying each trial's data with the precomputed common spatial filters.

Source space classification

We used LDA (see sensor space classification for details) to classify spatial frequency as well as image content from each frequency band and ROI. The virtual channel time series derived via LCMV beamforming served as features for the classifier. Statistical analysis of source space classification followed the same routine as described in the sensor space analysis.

Time frequency analysis

First, the evoked response was removed by subtracting the average ERF from each trial. Then, we calculated the time frequency representation of low and high temporal frequencies separately. For low frequency we extracted wavelets linearly space between 3-6 cycles for temporal frequencies ranging from 2-30 Hz in steps of one Hertz (time window -0.2-0.8 sec in steps of 0.05 sec). For the high frequencies, we used dpss multitapers (15 cycles of data at spectral smoothing of 0.2*frequency) for frequencies ranging between 35 and 125 Hertz in steps of 5 Hz. To determine individual gamma frequency, we compared high frequency power between baseline (-200 to 0 ms) and 100ms to 400ms after stimulus onset in the broadband source reconstructed ROIs using cluster-based permutation testing with dependent samples t-test on the trial-level over all conditions. Individual gamma was defined as the center frequency of the cluster with the maximum sum of t-values +/- 10 Hz. Participants where no significant cluster was found (3 out of 19 for Calcarine ROI) were assigned the average of all other participants (Fusiform ROI: 86.8113Hz Lateral Occipital ROI: 99.0183 Frontal ROI: 83.9073 Calcarine ROI: 90.6404). Alpha/beta ranges were derived from previous literature as 7-25 Hz (Michalareas et al., 2016) and split to 7Hz to 13Hz for alpha and 15Hz to 25Hz for beta. We calculated average power values for all frequency bands of interest by taking the average of the log transformed power values and baseline correcting to the -200 to -100 ms baseline window for each ROI and condition separately.

Gamma power comparison

We went on to test our hypothesis of reduced gamma in response to broadband images in the intact condition in EVC. First, we compared gamma power in the broadband versus the HSF condition separately for the intact and scrambled condition in EVC (separately for each time point). Then, we compared if gamma reduction in the broadband condition was stronger for the scrambled compared to the face condition. All tests were corrected for multiple comparisons using FDR corrections.

Cross frequency lagged correlations

We calculated lagged trial-by-trial pairwise correlation coefficients between the gamma power in the EVC ROI and alpha and beta power in all other ROIs using Spearman correlations with lags ranging from -150 to 150 in steps of 50ms. We did this separately for each condition and ROI for time points ranging from -0.1 to 0.7 sec. We tested if the correlation was different between the intact and scrambled conditions using cluster statistics (with dependent samples t-test as statistics, at an alpha of 0.05 with 1000 permutations).

References

Adrian, E. D. (1944). Brain rhythms. Nature Publishing Group.

Bar, M. (2004). Visual objects in context. *Nature Reviews. Neuroscience*, *5*(8), 617.

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., … Rosen, B. R. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(2), 449–454.

Barlow, H. (2001). Redundancy reduction revisited. *Network: Computation in Neural Systems*, *12*(3), 241–253.

Barlow, H. B. (1961). Possible principles underlying the transformation of sensory messages. *Sensory Communication*, *1*, 217–234.

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.

Belardinelli, P., Ortiz, E., Barnes, G., Noppeney, U., & Preissl, H. (2012). Source reconstruction accuracy of MEG and EEG Bayesian inversion approaches. *PloS One*, *7*(12), e51985.

Blais, C., Jack, R. E., Scheepers, C., Fiset, D., & Caldara, R. (2008). Culture shapes how we look at faces. *PloS One*, *3*(8).

Blankertz, B., Lemm, S., Treder, M., Haufe, S., & Müller, K.-R. (2011). Single-trial analysis and classification of ERP components—a tutorial. *NeuroImage*, *56*(2), 814–825.

Bullier, J. (2001). Integrated model of visual processing. *Brain Research Reviews*, *36*(2), 96–107.

Dalal, S., Jerbi, K., Bertrand, O., Adam, C., Ducorps, A., Schwartz, D., … Lachaux, J.-P. (2013). Simultaneous MEG-intracranial EEG: new insights into the ability of MEG to capture oscillatory modulations in the neocortex and the hippocampus. *Epilepsy and Behavior*, 10–1016.

Dalal, S. S., Rampp, S., Willomitzer, F., & Ettl, S. (2014). Consequences of EEG electrode position error on ultimate beamformer source reconstruction performance. *Frontiers in Neuroscience*, *8*, 42.

de-Wit, L. H., Kubilius, J., Wagemans, J., & de Beeck, H. P. O. (2012). Bistable Gestalts reduce activity in the whole of V1, not just the retinotopically

predicted parts. *Journal of Vision*, *12*(11), 12.

Fintzi, A. R., & Mahon, B. Z. (2014). A bimodal tuning curve for spatial frequency across left and right human orbital frontal cortex during object recognition. *Cerebral Cortex*, *24*(5), 1311–1318.

Fries, P. (2015). Rhythms for cognition: communication through coherence. *Neuron*, *88*(1), 220–235.

Friston, K., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1211–1221.

Gebodh, N., Vanegas, M. I., & Kelly, S. P. (2017). Effects of stimulus size and contrast on the initial primary visual cortical response in humans. *Brain Topography*, *30*(4), 450–460.

Goffaux, V., Peters, J., Haubrechts, J., Schiltz, C., Jansma, B., & Goebel, R. (2010). From coarse to fine? Spatial and temporal dynamics of cortical face processing. *Cerebral Cortex*, *21*(2), 467–476.

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., … Hämäläinen, M. S. (2014). MNE software for processing MEG and EEG data. *Neuroimage*, *86*, 446–460.

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., … Parkkonen, L. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, *7*, 267.

Gramfort, A., Papadopoulo, T., Olivi, E., & Clerc, M. (2010). OpenMEEG: opensource software for quasistatic bioelectromagnetics. *Biomedical Engineering Online*, *9*(1), 45.

Gregor, K., Szlam, A., & LeCun, Y. (2011). Structured sparse coding via lateral inhibition. *Advances in Neural Information Processing Systems*, *24*.

Groß, J., Kujala, J., Hämäläinen, M., Timmermann, L., Schnitzler, A., & Salmelin, R. (2001). Dynamic imaging of coherent sources: studying neural interactions in the human brain. *Proceedings of the National Academy of Sciences*, *98*(2), 694–699.

Hadjipapas, A., Lowet, E., Roberts, M. J., Peter, A., & De Weerd, P. (2015). Parametric variation of gamma frequency and power with luminance contrast: A comparative study of human MEG and monkey LFP and spike responses. *Neuroimage*, *112*, 327–340.

Hegdé, J. (2008). Time course of visual perception: coarse-to-fine processing and beyond. *Progress in Neurobiology*, *84*(4), 405–439.

Hillebrand, A., & Barnes, G. R. (2003). The use of anatomical constraints with MEG beamformers. *Neuroimage*, *20*(4), 2302–2313.

Hillebrand, A., & Barnes, G. R. (2011). Practical constraints on estimation of source extent with MEG beamformers. *Neuroimage*, *54*(4), 2732–2740.

Honey, C., Kirchner, H., & VanRullen, R. (2008). Faces in the cloud: Fourier power spectrum biases ultrarapid face detection. *Journal of Vision*, *8*(12), 9.

Hsieh, P.-J., Vul, E., & Kanwisher, N. (2010). Recognition alters the spatial pattern of FMRI activation in early retinotopic cortex. *Journal of Neurophysiology*, *103*(3), 1501–1507.

Janik, S. W., Wellens, A. R., Goldberg, M. L., & Dell'Osso, L. F. (1978). Eyes as the center of focus in the visual examination of human faces. *Perceptual and Motor Skills*, *47*(3), 857–858.

Kauffmann, L., Chauvin, A., Guyader, N., & Peyrin, C. (2015). Rapid scene categorization: Role of spatial frequency order, accumulation mode and luminance contrast. *Vision Research*, *107*, 49–57.

Kietzmann, T. C., Spoerer, C. J., Sörensen, L. K. A., Cichy, R. M., Hauk, O., & Kriegeskorte, N. (2019). Recurrence is required to capture the representational dynamics of the human visual system. *Proceedings of the National Academy of Sciences*, *116*(43), 21854–21863.

King, J. R., & Dehaene, S. (2014). Characterizing the dynamics of mental representations: the temporal generalization method. *Trends in Cognitive Sciences*, *18*(4), 203–210.

Kok, P., Jehee, J. F. M., & De Lange, F. P. (2012). Less is more: expectation sharpens representations in the primary visual cortex. *Neuron*, *75*(2), 265–270.

Kybic, J., Clerc, M., Abboud, T., Faugeras, O., Keriven, R., & Papadopoulo, T. (2005). A common formalism for the integral formulations of the forward EEG problem. *IEEE Transactions on Medical Imaging*, *24*(1), 12–28.

Ledoit, O., & Wolf, M. (2004). Honey, I shrunk the sample covariance matrix. *The Journal of Portfolio Management*, *30*(4), 110–119.

MaBouDi, H., Shimazaki, H., Amari, S., & Soltanian-Zadeh, H. (2016).

Representation of higher-order statistical structures in natural scenes via spatial phase distributions. *Vision Research*, *120*, 61–73.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190.

Marr, D. (1982). Vision: A computational investigation into the human representation and processing of visual information. *WH San Francisco: Freeman and Company*, *1*(2).

Mazer, J. A., Vinje, W. E., McDermott, J., Schiller, P. H., & Gallant, J. L. (2002). Spatial frequency and orientation tuning dynamics in area V1. *Proceedings of the National Academy of Sciences*, *99*(3), 1645–1650.

Michalareas, G., Vezoli, J., Van Pelt, S., Schoffelen, J.-M., Kennedy, H., & Fries, P. (2016). Alpha-beta and gamma rhythms subserve feedback and feedforward influences among human visual cortical areas. *Neuron*, *89*(2), 384–397.

Miellet, S., Caldara, R., & Schyns, P. G. (2011). Local Jekyll and global Hyde: the dual identity of face identification. *Psychological Science*, *22*(12), 1518–1526.

Norcia, A. M., Appelbaum, L. G., Ales, J. M., Cottereau, B. R., & Rossion, B. (2015). The steady-state visual evoked potential in vision research: a review. *Journal of Vision*, *15*(6), 4.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*, 1.

Parker, D. M., & Dutch, S. (1987). Perceptual latency and spatial frequency. *Vision Research*, *27*(8), 1279–1283.

Peli, E., McCormack, G., & Sokol, S. (1988). Signal to noise ratio considerations in the analysis of sweep visual-evoked potentials. *Applied Optics*, *27*(6), 1094–1098.

Penny, W. D., Friston, K. J., Ashburner, J. T., Kiebel, S. J., & Nichols, T. E. (2011). *Statistical parametric mapping: the analysis of functional brain images*. Academic press.

Perona, C. H. and P. (2000). Pasadena Houses.

Petras, K., ten Oever, S., Jacobs, C., & Goffaux, V. (2019). Coarse-to-fine information integration in human vision. *NeuroImage*, *186*, 103–112.

Popov, T., Oostenveld, R., & Schoffelen, J. M. (2018). FieldTrip Made Easy: An Analysis Protocol for Group Analysis of the Auditory Steady State Brain Response in Time, Frequency, and Space. *Frontiers in Neuroscience*, *12*.

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87.

Richter, C. G., Thompson, W. H., Bosman, C. A., & Fries, P. (2017). Top-down beta enhances bottom-up gamma. *Journal of Neuroscience*, *37*(28), 6698–6711.

Schwiedrzik, C. M., & Freiwald, W. A. (2017). High-level prediction signals in a low-level area of the macaque face-processing hierarchy. *Neuron*, *96*(1), 89–97.

Simoncelli, E. P. (2003). Vision and the statistics of the visual environment. *Current Opinion in Neurobiology*, *13*(2), 144–149.

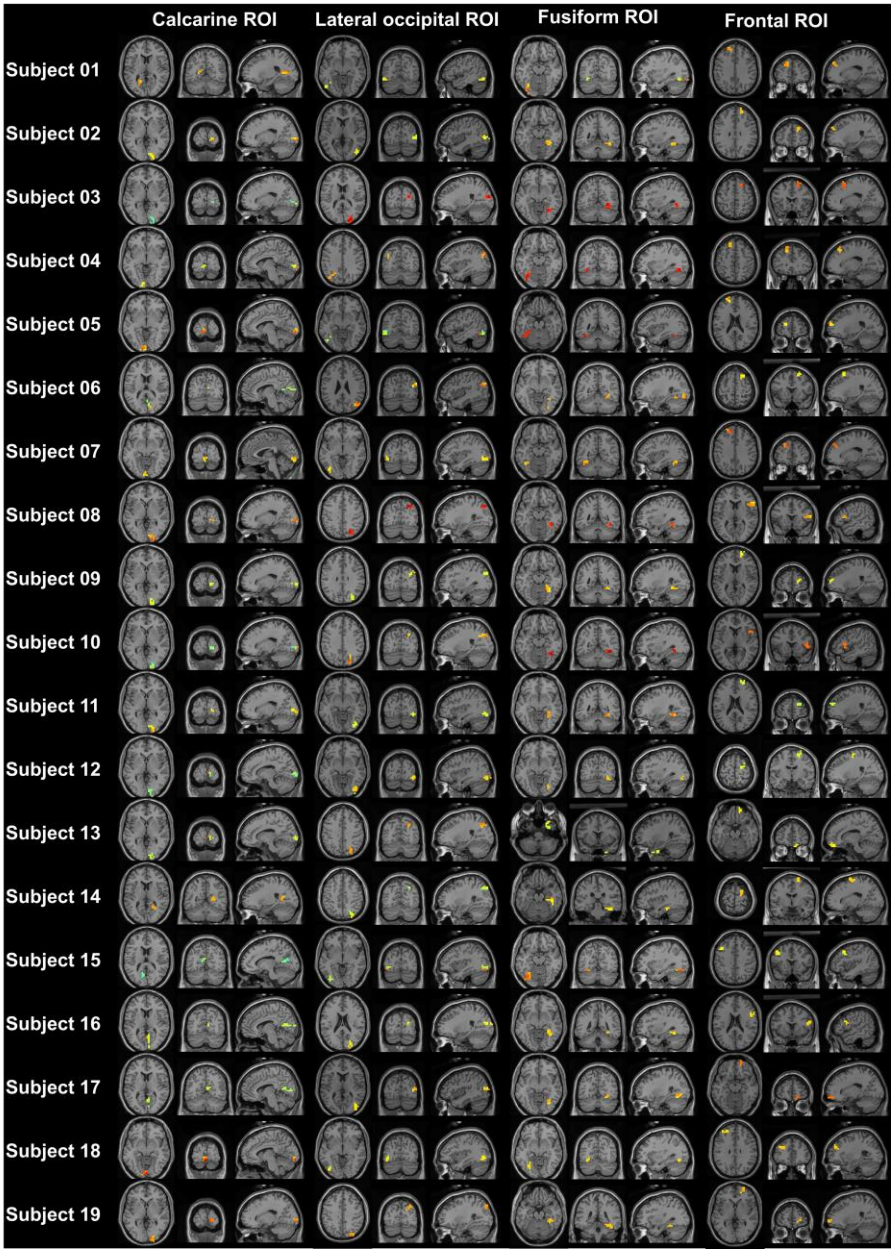
Strasburger, H., Scheidler, W., & Rentschler, I. (1988). Amplitude and phase characteristics of the steady-state visual evoked potential. *Applied Optics*, *27*(6), 1069–1088.

Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010). Controlling low-level image properties: the SHINE toolbox. *Behavior Research Methods*, *42*(3), 671–684.

Williams, M. A., Baker, C. I., De Beeck, H. P. O., Shim, W. M., Dang, S., Triantafyllou, C., & Kanwisher, N. (2008). Feedback of visual object information to foveal retinotopic cortex. *Nature Neuroscience*, *11*(12), 1439.

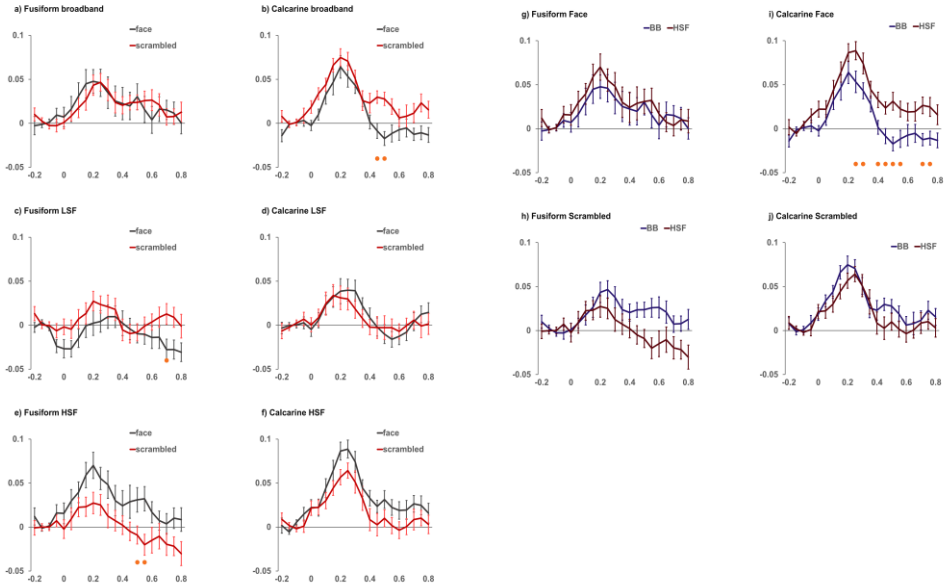
Supplementary Figures

1) ROIs of all participants

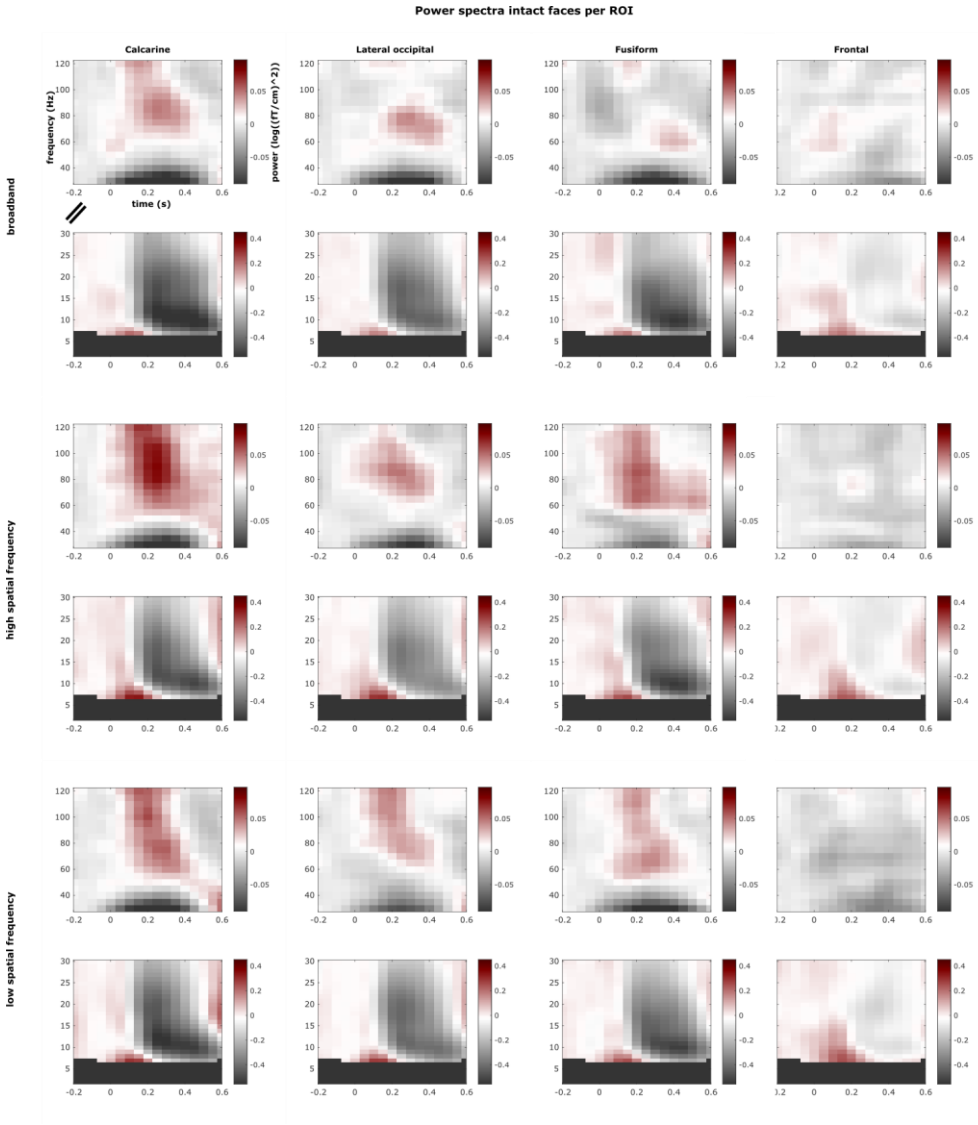


Supplementary figure 1: Voxels within the predefined anatomical ROI responding most strongly to faces for each participant.

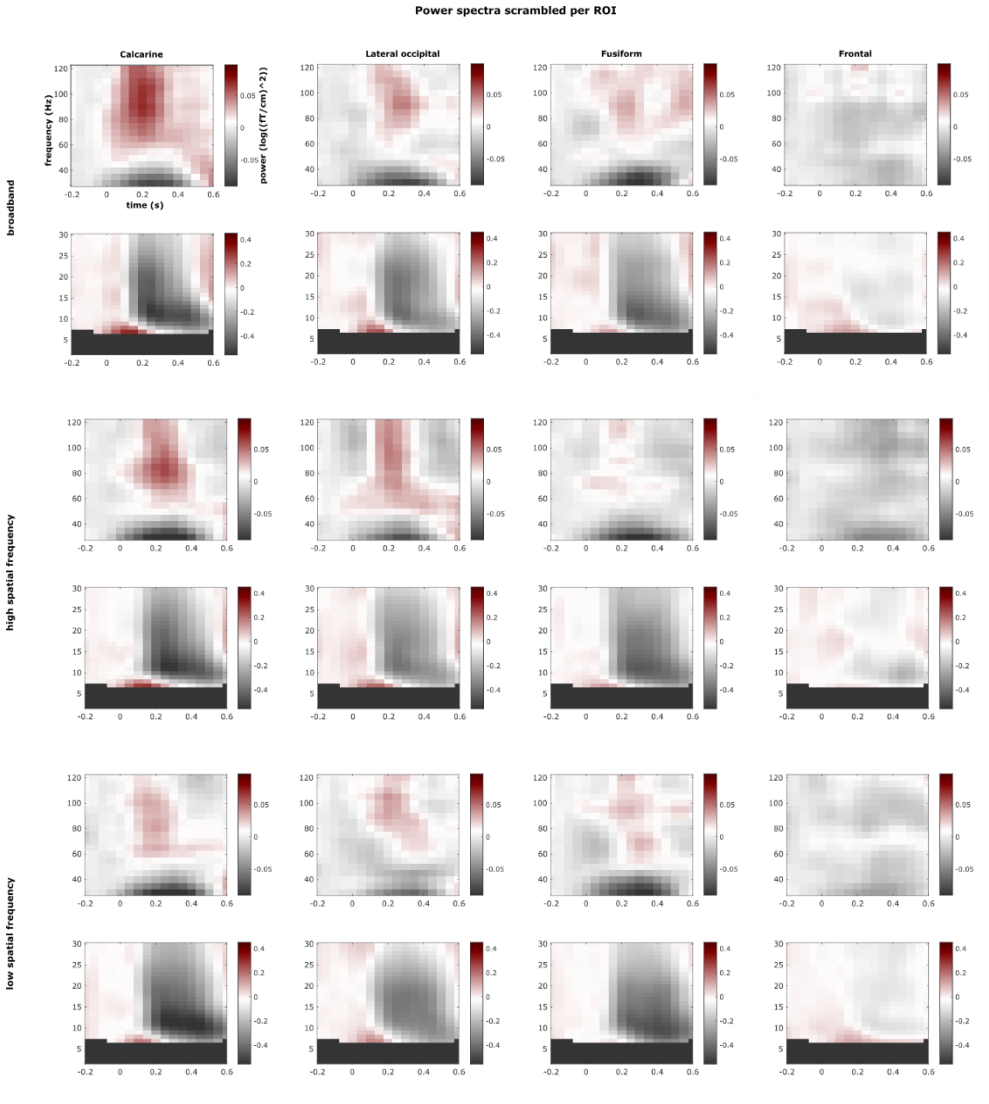
1) Gamma power as a function of latency



Supplementary figure 2: Gamma-range responses in the Fusiform and the Calcarine ROIs. Significant differences between conditions were found for intact face vs. scrambled face image trials in the Calcarine ROI when images contained broadband (BB) SFs with higher gamma power in response to scrambled images. Interestingly, the Fusiform ROI shows higher gamma power in response to scrambled, compared to intact face images when images contain only low spatial frequencies. When images contain high spatial frequencies (HSF), the expected pattern with higher gamma power for intact compared to scrambled images is found. Gamma power differed between broadband and HSF image trials only when faces were intact and only within the Calcarine ROI.



Supplementary figure 3a: Time-frequency tables of response to intact faces in all ROIs and spatial frequency conditions. Note that the frequency axes are not continuous between lower and upper frequency band panels.



Supplementary figure 3b: Time-frequency tables of response to scrambled faces in all ROIs and spatial frequency conditions. Note that the frequency axes are not continuous between lower and upper frequency band panels.

Chapter 5

Dissociating mid-level and high-level feedback to V1 during visual object perception

Kirsten Petras[1,2,*,†]; Sriranga Kashyap[2,*]; Jolien Schuurmans[1]; Christine Schiltz[3]; Benedikt A. Poser[2]; Valerie Goffaux[1,2,4]

1 Research Institute for Psychological Science, Université Catholique de Louvain, Louvain-la-Neuve, Belgium;
2 Department of Cognitive Neuroscience, Maastricht University, Maastricht, the Netherlands;
3 Institute of Neuroscience, Université Catholique de Louvain, Louvain-la-Neuve, Belgium;
4 Cognitive Science and Assessment Institute, University of Luxembourg, Luxembourg

* authors contributed equally
† corresponding author

Abstract

Visual processing of objects engages a distributed cortical network from the primary visual cortex to orbitofrontal regions and includes most of the ventral surface of the temporal lobes. Yet, how higher-level parts of this network contribute to the primary stages of processing is still unresolved. Competing theories make different claims about the origin and information content of feedback contributions to primary visual cortex with a focus either on areas within the occipital, or the temporal, parietal and frontal cortices. Here, we used sub-millimeter human fMRI recordings to investigate the origin and properties of feedback to V1 and V2 while participants viewed images of intact or distorted objects. We used two different image manipulations to disrupt object recognition or object coherence to dissociate between mid-level and high-level feedback influences. Preliminary data show, that we can resolve responses at a spatial scale suitable to distinguish feedforward from feedback dominated cortical layers.

Introduction

Human object recognition has been proposed to rely on feedback from higher-level category-selective visual regions that projects back to early visual cortex (EVC), including primary visual cortex (V1) (Bar, 2003; Bar et al., 2006; Kersten, Mamassian, & Yuille, 2004; Muckli et al., 2015; Williams et al., 2008). These feedback projections are thought to carry the result of a preliminary analysis of the visual input, to guide further processing in EVC.

Evidence from anatomical studies in animals however suggests that the majority of descending connections to V1 originate in low- to intermediate-level stages of the visual hierarchy, i.e. from regions tuned to relatively simple stimulus properties such as border ownership, boundaries, and contours. For example, Barone et al. (2000) used retrograde tracing in monkeys to rank visual regions based on the proportion of labelled cells located in supragranular layers of cortex (%SLN), where most feedforward connections originate (Felleman & Van, 1991; Markov et al., 2014; Scheeringa & Fries, 2019). Importantly, their data suggest that most feedback connections to V1 originate in hierarchically proximal areas. Similar findings are reported by Markov et al. (2014) who further determined that laminar layer 2/3A is a short distance feedback stream, receiving input from neighboring areas, while layer 5/6 receives longer distance feedback connections from areas further downstream, as well as some proximal feedback. Overall, the authors found that 75% of corticocortical neurons in inter-areal connections are short to medium range, with comparatively few distal feedback connections.

To reconcile the proposed importance of feedback to V1 from hierarchically distant category-selective visual areas with the low number of such projections relative to those from hierarchically more proximal regions, we consider three possible scenarios (see figure 1 for overview):

In **scenario one**, feedback from category-selective cortex is efficiently transmitted to V1 such that few connections have a large impact on V1 processing. This scenario is supported by the drastic increase in receptive field (RF) sizes along the ventral visual pathway. For example, in macaque monkeys, Shushruth et al. (2009) found V1 receptive field sizes around 0.3° ± 0.13° while Op de Beek & Vogels (2000) found RFs in anterior portions of inferior temporal (IT) cortex covering up to 26° of visual angle, with most neurons responding to fields of about 12° in size. Kobatake & Tanaka (1994) found a stepwise increase of RF size along the visual hierarchy with small RFs in V2 (1.7±1, mean ± SD), intermediate RF sizes in V4 (4.8±2.5) and posterior IT (5.4±2.8) and large RFs in anterior IT (16.5±6.1). Other studies have shown that feedback to V1 is not retinotopically restricted to the cortical patches relaying the initial feedforward input but acts more globally (de-Wit, Kubilius, Wagemans, & de Beeck, 2012; Williams et al., 2008). Thus, a divergence of few feedback connections onto many V1 receptive fields seems plausible.

In **scenario two**, feedback modulations that are usually associated with higher level visual areas are instead largely driven by basic image properties that can be resolved in mid-level visual regions, such as figure/ground segregation, curvature and contours as well as histogram statistics (Kubilius, Wagemans, & Op de Beeck, 2014). Numerous studies find strong relationships between the statistical properties of natural images and their category membership both in the images themselves (e.g. Torralba & Oliva, 2003) and in the neural responses they evoke (Kersten et al., 2004; Kersten & Yuille, 2003; Rossion, Hanseeuw, & Dricot, 2012). Thus, correlations of higher-level visual regions with V1 feedback modulations might be erroneously attributed to causal influences of said areas and instead be the result of intermediate level areas influencing both V1 and higher-level processing.
Finally, in **scenario three**, some proportion of feedback generated in high-level areas could cascade down through intermediate levels of the hierarchy, rather than projecting directly from distal areas into V1. This notion is in line with the reverse hierarchy theory (Hochstein & Ahissar, 2002; Juan & Walsh, 2003), which purports that V1 receives feedback from higher-level cortical areas along the classical visual hierarchy in reverse order.

To disambiguate feedback origins, we propose to combine an imaging technique that enables separating feedforward and feedback signals with stimulation protocols that detach mid-level from high-level processing.
Methodological advancements in ultra-high field (UHF) fMRI now allow measurement of signals from voxels that are much smaller (sub-millimeter scale) than the average gray matter thickness of V1 with sufficiently high signal-to-noise ratio (Goense, Merkle, & Logothetis, 2012; Kashyap et al., 2018; Koopmans, Barth, Orzada, & Norris, 2011; Polimeni, Fischl, Greve, & Wald, 2010). Given that feedback signals predominantly target deep and superficial layers, whereas feedforward signals predominantly target the middle (granular) layer (Harris & Mrsic-Flogel, 2013; Markov et al., 2014), UHF fMRI at sub-millimeter spatial resolutions can be used in human participants to distinguish feedforward from feedback dominated signals across cortical depths. Previous laminar fMRI studies have used paradigms that prevented feedforward signals from reaching the neural populations under investigation to isolate top-down effects. Kok et al. (2016) found illusory contours from Kanizsa figures (Kanizsa, 1976) to selectively increase BOLD responses in deep layers of V1, consistent with feedback from distal regions (Markov et al., 2014). Muckli et al. (2015) used a paradigm were participants viewed natural scene images, while the image quadrant that corresponded to the retinotopic patch of cortex under investigation was occluded to prevent feedforward signals from entering their region of interest. They found that the activation in superficial layers in retinotopic patches representing the occluded area was predictive of the scene

outside the occluded area. Both studies found evidence of feedback to V1. However, their designs are only suitable to determine the minimal hierarchical level necessary for its generation. For instance, the observed effects in Kok et al.'s study, require feedback to be generated in areas that respond to illusory contours, so minimally V2 (Von der Heydt, Peterhans, & Baumgartner, 1984), but multiple more distal regions beyond V2 could be equally suitable sources.

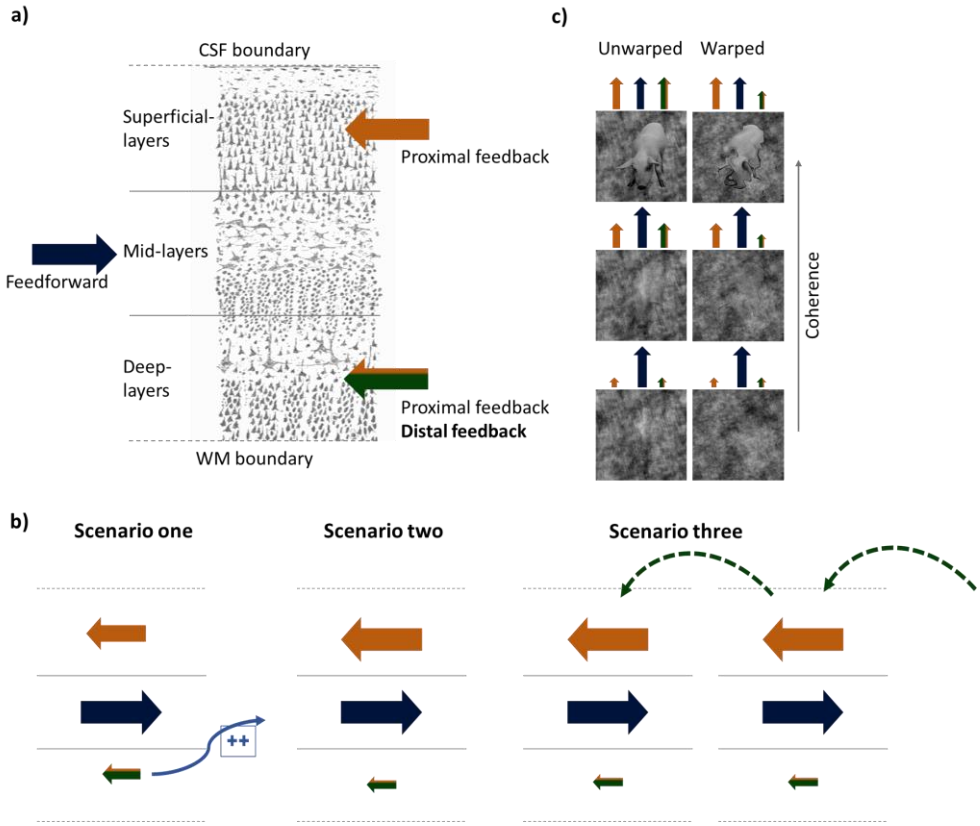


Figure 1: a) Supragranular (superficial) layers receive mostly feedback from proximal areas. Granular layer IV receives feedforward connections and falls within the mid-layer bin and infragranular, deep layers receive mostly distal, but also some proximal feedback. Note that the cortical depth bins that can be resolved with fMRI do not exactly correspond to histological layers as shown here with the superimposed scaled example of a drawing by Ramón y Cajal. b) In scenario one few distal feedback connections exert a large influence over feedforward output (light blue arrow). In scenario two most feedback comes from proximal areas and distal feedback is negligible and in scenario three feedback originates in high level, distal areas, but cascades down via proximal connections. c) The two different image manipulations are expected to differentially affect signaling to different cortical

depth. While warping is expected to disrupt mostly the distal feedback to deep layers, loss of phase coherence is expected to disrupt both, superficial and deep layer targeting feedback. Based on previous literature, we further expect an increase in mid-layer processing with decreasing phase coherence.

In order to investigate the hierarchical level of feedback sources with greater specificity, here, we manipulate images of animals, plants and manmade objects to disrupt their processing mainly in high-level category-selective visual regions, or in both high- and mid-level visual regions, while we target our analysis on V1 and V2. By parceling out the contribution of high- versus mid-level regions to the overall activity in feedback dominated cortical layers, we derive a measure of their unique contribution to feedback activity.

We acquired functional images at an isotropic voxel size of 0.8mm while participants viewed images of a diverse set of natural and manmade objects that were either left undistorted, or warped to disrupt image recognizability (Stojanoski & Cusack, 2014a, also see methods for details). In addition, images were presented at three levels of Fourier phase coherence ranging from fully coherent to fully incoherent (i.e. scrambled) with a third coherence level just below the threshold at which participants were able recognize the object (see methods and pilot experiment for details). Image warping, where a coherent shape is still present, but the visual object is made unrecognizable through distortion, is thought to mostly affect higher-level visual areas (above V4) while preserving response properties in early visual cortex. Fourier phase scrambling, where shape is lost through the disruption of phase alignment, also renders images unrecognizable. Compared to warping however, phase scrambling has been suggested to disrupt processing earlier in the visual hierarchy (Stojanoski & Cusack, 2014b). Thus, while both warping and phase scrambling should affect higher-level category-selective cortex, scrambling should have a stronger effect on lower- to intermediate-level visual areas. Because phase scrambling removes all shape and meaning from the image, we expect fully phase-scrambled (i.e. 0% phase coherence) images of both the unwarped and the warped type to evoke similar responses at cortical depths associated with feedback processing (further referred to as *FB layers*) as well as at depths associated with feedforward processing (further referred to as *FF layers*). When images are coherent in phase (100% coherence), we expect unwarped images to engage high-level category-selective cortical areas to a greater extent than warped images with no information on object identity, while both image types engage mid-level visual regions similarly. Finally, we expect partial phase coherence (perceptual recognition threshold, see pilot experiments) to affect the processing of unwarped images more than that of warped images as in warped images, object meaning is already lost.

In this paradigm, each of the above-mentioned scenarios is associated with a specific set of testable predictions.
Under the assumption of **scenario one** (few feedback connections with large effect) we expect activation differences in long distance FB layers of V1 for undistorted, compared to warped images. Concretely, we expect unwarped images to strongly engage visual areas beyond V4 while warped images should mostly be limited to engage visual regions up to V4. Consequently, unwarped images are expected to generate high-level feedback to V1. In response to warped images, the difference in response between fully and partially coherent images is reduced in V1 feedback. Modulations of FB layers of V1 should be reflected in the FF layer of V2, that receives much of its input from V1. When images are fully incoherent, we expect similar responses to unwarped and warped images in all layers of V1 and V2.

In **scenario two** (mid-level feedback), most feedback comes from mid-level visual areas that are similarly activated by unwarped and warped images, that both preserve basic image properties such as a coherent shape and clear figure/ground boundaries. In this scenario, we expect similar feedback behavior for both image types. Concretely, we expect feedback to be mostly affected by coherence, and less by warping type. In effect, FB layer activations are expected to change across coherence levels in a similar way for unwarped and warped images.
If instead **scenario three** (reverse hierarchy) best describes feedback organization, we expect the effect of the different coherence levels to differ between intact and warped images, with a steeper decrease of activation in FB layers for intact images, than for warped images that contain no category information at any level of coherence.. We further expect the rate of FB decrease in cortical depths closest to short range feedback layers 2/3a to be correlated with V2 *FF*.

Methods

Participants

3 healthy volunteers with normal or corrected to normal visual acuity participated in the MRI experiment. They were reimbursed for their time with 10euros per hour in the form of gift vouchers. MRI participants underwent two scanning sessions on separate days that lasted 2 hours each. All MRI procedures were approved by the Ethics committee of the Faculty of Psychology and Neurosciences at Maastricht University. A total of 117 volunteers participated in two behavioral pilot experiments. The pilot experiments were conducted using an online testing platform (testable.org). Online participants entered a raffle for 5x20 euros and. Online participants could only participate in one of the two experiments.

All participants provided written or digital informed consent.

Stimuli

Stimuli were created from initially 143 high resolution (min 800x800 pixels) full color photographs of single objects, animals and plants (81 natural and 62 manmade). All images were reduced to 800x800 pixels with the object in the center, cropped from their background where necessary and mirror padded to 1200x1200 pixels. We used two types of transformations to render images unrecognizable: Fourier phase scrambling and diffeomorphic warping.

In Fourier phase scrambling the amplitude spectrum of the image is preserved so that spatial frequency and orientation content are closely matching those of the original image while the phase information is manipulated (e.g. Goffaux et al., 2010; Tsao, Freiwald, Knutsen, Mandeville, & Tootell, 2003). To allow for partial scrambling while avoiding artifacts due to non-uniform phase steps that can arise when not taking the circular distribution of phase values into account, we chose an approach suggested by Ales et al (2012). We linearly interpolated between phase angles in the direction of the minimum circular distance between them therewith ensuring a uniform distribution of phase angles around the unit circle. To mitigate the effect of the uniform, low spatial frequency background leaking into the image area and reducing the spatial frequency, we iteratively re-applied the original object onto the phase-scrambled background and re-scramble (Petras, ten Oever, Jacobs, & Goffaux, 2019). We used 1000 iterations per image.

As a second manipulation we used diffeomorphic transformations. For a detailed explanation of the warping procedure see Stojanoski & Cusack (2014). In brief, diffeomorphic transformations render images unrecognizable while preserving image properties such as shape size and local distribution of contrast by stretching and compressing the original image in randomized directions. Images were up-sampled by a factor of two to reduce blurring due to interpolation errors caused by warping and the background color was set to green in order to aid the creation of background masks. To produce the distorted images a flow field was created from random phase and amplitude cosine components in two dimensions and repeatedly applied to the image. Because every pixel in the original image had an equal chance of being expanded or contracted in the warped image, the average number of object and background pixels is equal between intact and warped images. Further, because the warping procedure does not introduce any discontinuities in the pictured object, the number of connected object pixel clusters is the same as in the original image. In contrast to Fourier phase scrambling or box scrambling, diffeomorphic transformations have been demonstrated to be indistinguishable from non-distorted images within all 3 layers of the HMAX model of early visual object recognition (Stojanoski & Cusack, 2014b).

To ensure warped images, as well as partially phase-scrambled images were not recognized by participants, we conducted two behavioral pilot experiments to titrate our image set.

For pilot experiment 1, we diffeomorphically warped images in 8 steps centered on the averaged perceptual rating from 415 participants reported by Stojanoski and Cusack (2014). We excluded the ratings of warped faces and bikes because those categories seemed less susceptible to scrambling in the original study. We included no faces or bikes in our stimulus set for the same reason. To determine the level of warping necessary for the disruption of object recognition we presented 101 participants with sequences of progressively less distorted images. Each image was presented for one second (see figure 2a for an example sequence). Participants indicated when they recognized the object in the image by pressing a button and were then asked to select the correct object category from 5 options. For each individual image, we selected the warping level that corresponded to the image in the sequence two seconds before the mean recognition response of all participants. Four images (all from the manmade group) from the original set were excluded from the main experiment because the perceptual ratings were inconclusive.

For pilot experiment 2 we Fourier phase-scrambled images between 30 and 90% phase coherence in 20 steps. Sixteen participants indicated via a button press when they first recognized the object in the images presented in a sequence of ascending coherence (see figure 2b for example sequence). Each image was on screen for one second. For each individual image, we selected the coherence level that corresponded to the image in the sequence one second before the mean response time of all participants.

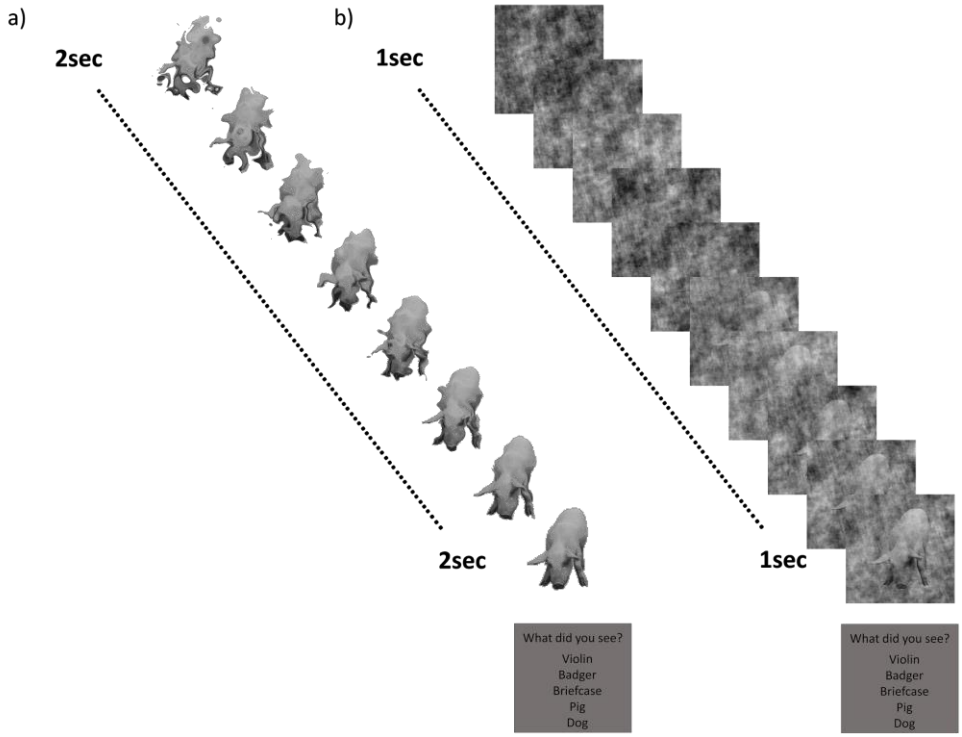


Figure 2: a) Images were warped in 8 steps. Each image remained on screen for 2 seconds before being replaced with the next lower warping step until the participant indicated recognizing the object in the image. b) Images were phase-scrambled in 20 steps (only every other step is shown here). Each image remained on screen for one second before being replaced with the next coherence step until the participant indicated recognizing the object in the image. Note the presence of an object becomes apparent before the object can be identified.

Stimuli in the main experiment were 81 natural and 58 manmade objects manipulated to 2 warping levels (no warping or warping at the level determined in pilot experiment one) and three phase coherence levels (100% phase coherence, 0% phase coherence, phase coherence determined by the level determined in pilot experiment 2). Images were projected onto a screen and viewed via a mirror attached to the head-coil at a viewing distance of 99 cm resulting in an image coverage of 5 degrees of visual angle. For example-stimuli of all conditions see figure 3.

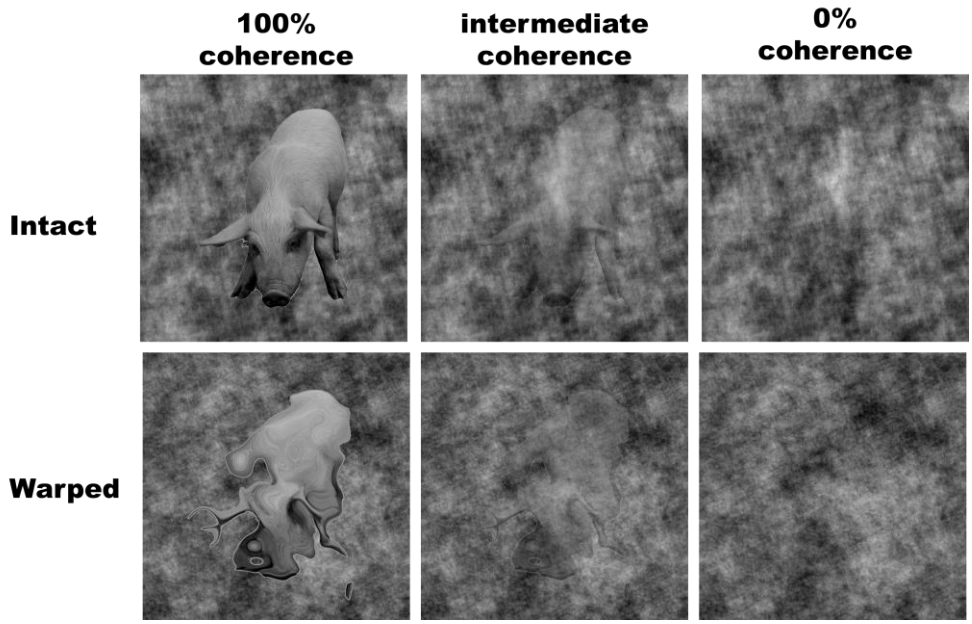


Figure 3: Example stimuli. Images were either non-warped or warped and were phase-scrambled to preserve either 100% phase coherence, intermediate phase coherence or no phase coherence. The level of warping necessary to disrupt recognition as well as the level of phase coherence at the perceptual threshold were titrated in separate experiments.

To estimate population receptive fields (pRF, see Dumoulin & Wandell (2008) for details), we presented retinotopic mapping stimuli consisting of oriented bars in four orientations and eight positions per orientation, resulting in 32 configurations. Stimuli and code to run and analyze pRF mapping can be found here: https://github.com/ingo-m/pyprf.

Procedure

Every participant completed 10 functional runs over two sessions. Each functional run consisted of 36 blocks lasting 4 seconds each with a 12 second fixed fixation period in between blocks. Conditions were counterbalanced across blocks. A block consisted of 10 images of different objects in one of the distortion conditions that were interleaved with 10 blank images (see figure 4b). The visual stimulation area was flanked by two vertical black bars. The participant was asked to press a button whenever both bars simultaneously changed color to white. Each run contained 12 targets.

In addition, each participant completed at least 2 pRF runs of 9min each. During pRF runs, participants were instructed to press a button whenever the central fixation dot changed color.

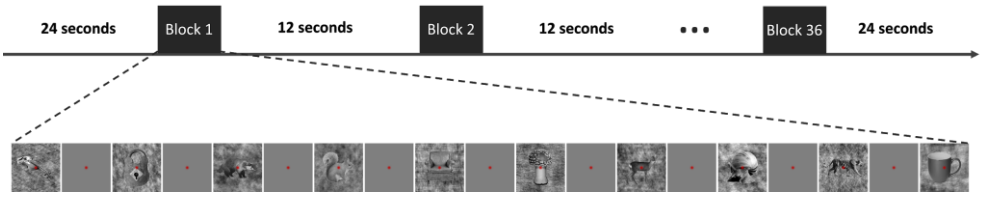


Figure 4: Each run consisted of 36 blocks lasting 4 seconds each separated by a 12 second inter block interval. Each block consisted of 10 images interleaved with blank intervals. The order of blocks was counterbalanced between runs and participants and blocks of each condition had an equal chance to contain a target.

MRI acquisition

All data were acquired on a whole-body Siemens Magnetom 7T research scanner (Siemens Healthineers, Erlangen, Germany) using a 32-channel whole-brain coil (Nova Medical, USA) at Scannexus B. V., Maastricht, the Netherlands. The MRI acquisition consisted of two sessions per participant and at least five functional runs were acquired in each session. The study was approved by the Ethics Review Committee for Psychology and Neuroscience (ERCPN) at Maastricht University (approval number ERCPN-OZL 200_04_10_2018) and all procedures followed the principles expressed in the Declaration of Helsinki.

Session 1

Anatomical data was acquired at 0.8mm isotropic resolution using a 3D-MP2RAGE sequence (Marques et al. 2010) (208 sagittal slices; TR=5000 ms; TE=2.43 ms; TI1/TI2=900/2750 ms; α1/α2=5°/3°; PF_{phase}=7/8; GRAPPA=3; Ref. lines PE=24; BW=250 Hz/px; echo-spacing=6.7 ms; TA=7:32 min). At least five functional runs were acquired at 0.8mm isotropic resolution using a gradient-echo 3D-EPI sequence (Poser et al. 2010) (40 oblique, coronal slices; $TR_{effective}$=2046 ms; TE=25 ms; phase-encoding=R>L; α=15°; PF_{phase}=6/8; PF_{slice}=6/8; GRAPPA=3; Ref. lines PE=78; BW=1108 Hz/px; echo-spacing=1.01ms; BWTP=25; PAT refscan=Flash; TA=10:50 min). Population receptive field (pRF) mapping data was acquired at 1mm isotropic resolution using a gradient-echo 3D-EPI sequence (40 oblique, coronal slices; $TR_{effective}$=2046 ms; TE=25 ms; phase-encoding=R>L; α=15°; PF_{phase}=6/8;

PF_{slice}=6/8; GRAPPA=3; Ref. lines PE=78; BW=1112 Hz/px; echo-spacing=1ms; BWTP=25; PAT refscan=Flash; TA=8 min). Five volumes of opposite-phase encoded data were acquired after every 3D-EPI run for distortion-correction.

Session 2

A 1mm isotropic resolution 3D-MP2RAGE sequence was acquired to aid highly accurate co-registration between sessions. All functional runs in this session were acquired at 0.8 mm isotropic resolution using the gradient-echo 3D-EPI sequence with the same parameters as in session 1.
A total of ten functional runs and 2 pRF mapping runs were acquired in both sessions combined.

MRI data analysis

All data were first converted to the BIDS format (Gorgolewski et al., 2016) to facilitate data sharing.
Anatomical:
The session 1 MP2RAGE was pre-processed in SPM12. The pre-processing steps included bias-field correction, skull stripping, and cleaning up of the dura, sagittal and transverse sinuses (https://gitlab.com/skash/sagittal_vein_masking).
In subjects where the automatic workflow performed sub-optimally, manual clean-up was carried out using ITK-SNAP v3.8.0. The pre-processed T_1-w dataset was segmented using the high-resolution recon-all pipeline in Freesurfer v6.0 at the native resolution. The MP2RAGE T_1 map was used as a proxy for T_2 in -*autorecon3* due to their similar contrasts in order to further optimize accurate pial boundary placement. The MP2RAGE from session 2 was up-sampled using Sinc interpolation to 0.8mm isotropic using the "*ResampleImage" function* in ANTs (https://github.com/ANTsX/ANTs). The up-sampled MP2RAGE from session 2 was co-registered to the session 1 MP2RAGE using *antsRegistration* and the inter-session transformations were saved. The Session 1 MP2RAGE and the segmentation masks from Freesurfer (https://surfer.nmr.mgh.harvard.edu/) were upsampled to 0.4mm isotropic and small errors due to upsampling in the GM and WM segmentation in the ROI were manually corrected using ITK-SNAP.
Functional: The functional runs in each session were pre-processed separately using ANTs unless mentioned otherwise. The realignment of the functional time-series was carried out similar to the SPM12 two-pass procedure but scripted in ANTs using the *antsRegistration*. The mean of each realigned run and the mean of the realigned opposite phase-encoded data were used to estimate the distortions. The anatomical data in session 1 and session 2 were co-registered to the realigned, unwarped functional data and the transformation matrices were saved. Please note the transformation matrices in ANTs can be applied forward and inverse using

antsApplyTransforms. The transformation estimated between the session 2 resampled anatomical and the session 1 anatomical due to their superior image contrast were concatenated together with the realignment and distortion-correction transforms and applied in a single resampling step using *antsApplyTransforms* for all functional data. The functional data from session 2 included the inter-session coregistration transform in addition to the realignment and distortion-correction. This multi-stage process allowed accurate transformation of all functional runs into a common space without multiple resampling steps whilst ensuring optimal within-session motion and distortion-correction.

All the pre-processed functional data were further analyzed using FSL's FEAT v6.0. The data were high-pass filtered using a Gaussian-weighted least-squares fitting (σ=100s). Statistical analysis was carried out using FILM with local autocorrelation correction (Woolrich et al., 2001) The time-series model included a standard two-gamma hemodynamic response function with temporal derivatives. The within-subject higher-level analysis was carried out using a fixed effects model by forcing random effects variance to zero in FLAME (Beckmann et al., 2003, Woolrich et al. 2004). No spatial smoothing was applied at any stage of the processing or analysis.

pRF: The pre-processing of the pRF runs were similar to the session 1 functional data and included realignment and distortion-correction. The pRF fitting was carried out using PyPRF (https://github.com/ingo-m/pyprf). The pRF metrics obtained from the fitting procedure were then resampled using nearest neighbor interpolation onto the Freesurfer cortical mesh reconstruction for surface-based analyses. In addition to our pRF maps, we used *neuropythy* to apply the Wang atlas (Benson, Butt, Brainard, & Aguirre, 2014; Wang, Mruczek, Arcaro, & Kastner, 2015) to subject space as another reference for the delineation of visual regions V1 and V2 (https://github.com/noahbenson/neuropythy). The spatial extent of the ROI was defined as the bounds of the stimulus by excluding the edge region (0.25^{o}) of the stimulus on each side. These spatial extents were thresholded using the R2 estimates (>0.1) from the pRF fitting (Marquardt et al. 2018) and were separately combined with V1 and V2 masks to obtain our sampling ROIs.

Extraction of laminar profiles

The 0.4 mm isotropic GM and WM masks were first converted to the levelset representation using CBS-Tools (Bazin et al., 2014) and 12 equivolume layers (Waehnert et al., 2014) were estimated. The laminar profiles were obtained from the V1 and V2 ROIs. The finely sampled laminar signals were then transformed into three bins representing: infra-granular, granular and supra-granular layers.

Statistical analysis

Parameter estimates produced by the fixed effects model (see above) are converted to Z-scores and all contrasts are evaluated at the single participant level.

Results

Behavioral pilot experiments for stimulus adjustment

We used diffeomorphic transformations as well as Fourier phase scrambling to manipulate the recognizability of naturalistic images of animals, plants and manmade objects. To determine the respective levels of distortion and phase scrambling needed to disrupt image recognition we conducted two behavioral pilot experiments. In the first pilot experiment participants saw sequences of progressively less distorted images of the same object and responded with a button press as soon as they recognized the pictured object. In the second pilot a separate group of participants saw sequences of images that progressively increased in phase coherence and responded with a button press as soon as they recognized the pictured object. Figure 5 shows the average distribution of participant responses for each of the pilot experiments. We analyzed response times separately for each image sequence to account for the variability in individual images susceptibility to both warping and phase incoherence. In the main experiment each image was assigned the warping level corresponding to the level on screen two seconds before the mean correct response time. As for coherence, each image was assigned the level corresponding to the level on screen one second before the mean correct response time. The average warping step across all images was 5 and the average coherence was 33.31%.

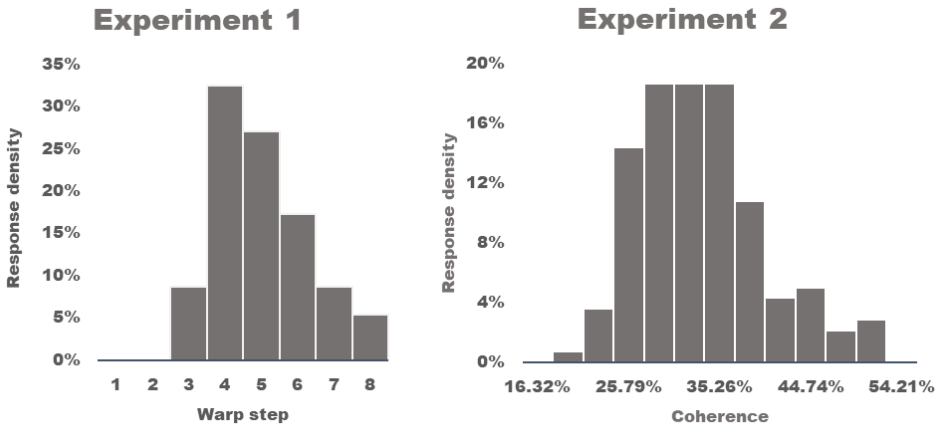


Figure 5: Response densities over warp step (experiment 1) and coherence (experiment 2) across all participants and images.

Data quality and retinotopic mapping

Thermal noise, i.e. the largely Gaussian-distributed additive fluctuations in the signal that are generated by thermal fluctuations in the receiver electronics as well as in the participant's tissue, are exacerbated when voxels are small, because thermal SNR scales linearly with voxel size (Brooks, Faull, Pattinson, & Jenkinson, 2013). The 3D EPI sequence we used to acquire functional data has been shown to be advantageous in thermal noise dominated sub-millimeter voxel regimes (Huber et al., 2018). We assessed the quality of our functional data by comparing the temporal signal-to-noise ratio (tSNR) across runs and sessions. In accordance with the measurements made by Huber et al. (2018) with a similar sequence, we achieved tSNRs up to 35 (see figure 6).

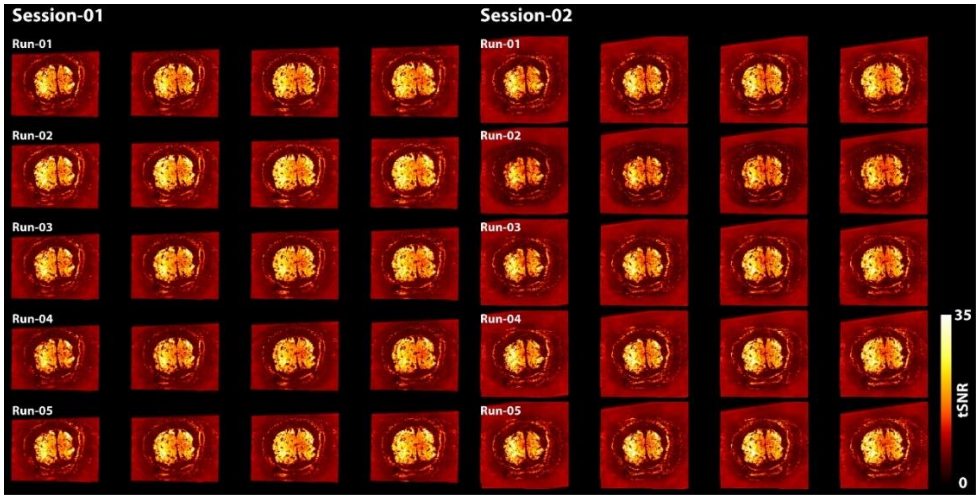


Figure 6: Run-wise tSNR maps for each session for one representative participant. Four example coronal slices from the occipital cortex are shown to illustrate both, the temporal stability of the timeseries data and the quality of our inter-session pre-processing workflow.

To delineate primary and secondary visual cortex, we used pRF mapping and confirmed our delineations with the help of Wang (2015) probabilistic maps. Figure 7 shows the pRF derived eccentricity and polar angle maps of one representative participant along with the delineation of V1 and V2 on an inflated sphere.

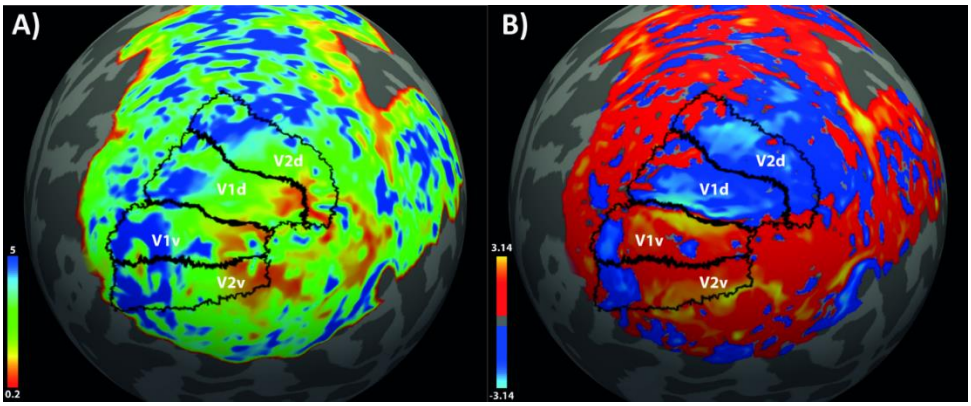


Figure 7: Eccentricity (A) and polar angle maps (B) of one example participant. Both maps are derived from two runs of PRF mapping. In (A), cooler colors reflect larger eccentricity. The delineation of V1 and V2 based on Wang et al. (2015) is superimposed and maps are inflated to a sphere. Lower case d indicates dorsal and lowercase v ventral aspects of V1 and V2. Note in (B), that the Wang probabilistic map based delineations of the ventral/dorsal V1 boundary closely matches the pRF derived polar angle map (reversal of polar angle at the boundary).

For the cortical depth dependent analysis, we defined layers such that total volume was equal between depth – bins (equi-volume sampling, see Waehnert (2014)). Figure 8 shows the layer profile of an example participant.

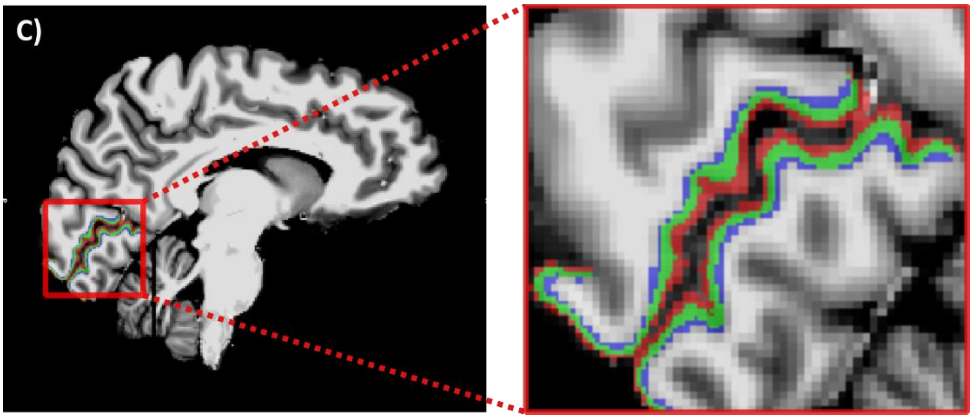


Figure 8: Illustation of the 3 laminar bins corresponding to infragranular (purple), granular (green) and supragranular (red) layers in V1 of an example participant

Cortical depth dependent activation

Preliminary analysis shows that our data produces the characteristic laminar profiles, with increases in parameter estimates with reduced cortical depth (i.e. towards the CSF boundary). Figure 9 illustrates the average Z-scores over all conditions for one example participant and figure 10 shows the individual condition z-scores.

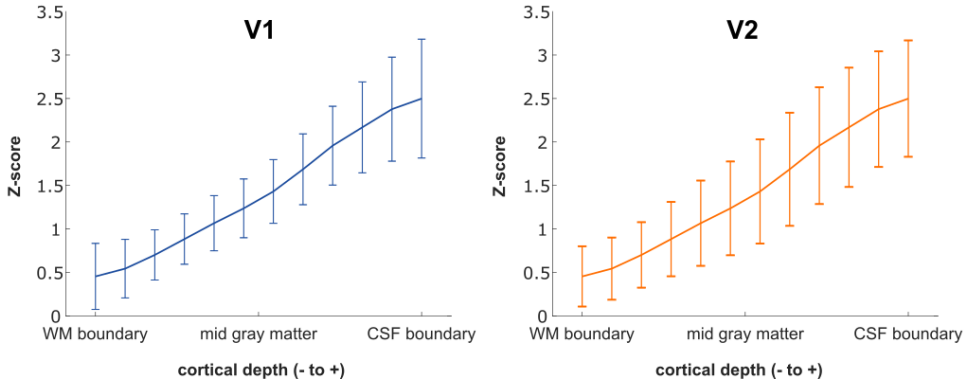


Figure 9: Laminar profile of responses averaged over all conditions for one example participant. In both V1 (left) and V2 (right) parameter estimates increase towards superficial layers.

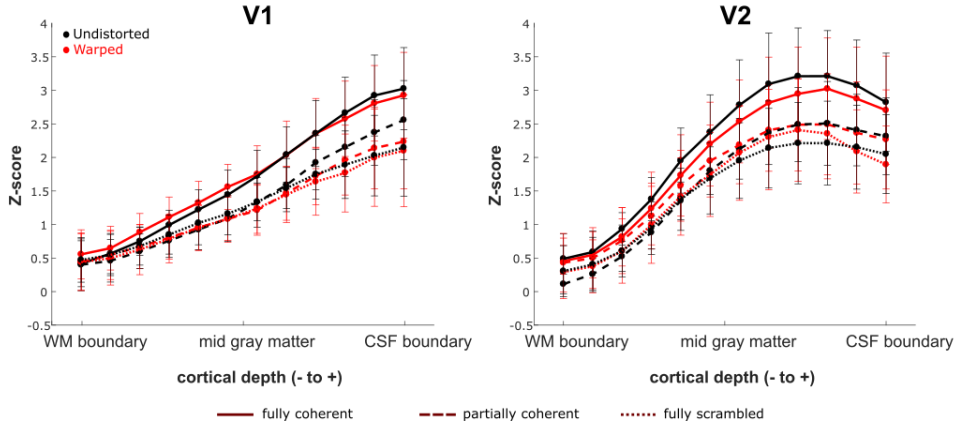


Figure 10: Laminar profile of responses to each condition for one example participant. Error bars reflect the standard deviation across runs.

Discussion

We used UHF fMRI to investigate the mechanics of feedback to V1 during visual object processing. We identified three possible scenarios and will discuss evidence for and against each of them here.

So far, most studies investigating feedback signals from extrastriate visual areas to V1 used simple stimuli that are likely resolved at early to intermediate levels of the visual hierarchy (Hupé et al., 1998; Marquardt, Schneider, Gulban, Ivanov, & Uludağ, 2018; Poort, Self, Van Vugt, Malkki, & Roelfsema, 2016). Our stimuli were designed to either selectively disrupt feedback from high-level ventral visual areas (warping manipulation), or to disrupt all object processing more globally (coherence manipulation). Evidence from two recent MEG studies suggests that feedback influences during visual object processing originate in hierarchically higher, category-selective areas in the ventral visual pathway. Michalareas et al. (2016) observed Granger causal influences of areas in the ventral visual cortices on V1 processing. Kietzmann et al. (2019) find areas beyond V3 to contribute unique feedback information to early visual cortex. However, the results of both studies can be explained equally well in both scenario one (few high-level feedback connections with large effect) and scenario three (reverse hierarchy).

Using cortical depth resolved fMRI, we can differentiate between proximal and distal feedback based on the different laminae they target. We will collect and analyze more data on this project before drawing conclusions with regard to our hypotheses.

So far, preliminary analysis shows, that the tSNR our measurements provided is suitable to conduct cortical depth dependent analysis. The initial laminar profiles with an increase in parameter estimates toward superficial layers are in line with previous reports on sub-millimeter data (Havlicek & Uludağ, 2020; Marquardt et al., 2018).

Future analysis of our data will also entail a comparison between stimulated and non-stimulated parts of V1 in response to intact and warped images. Previous research has found feedback to affect areas of V1 that did not correspond to the retinotopic locations of stimulation (Williams et al., 2008). If we find condition differences in superficial and deep, but not mid gray matter cortical depth in non-stimulated patches of V1, we will interpret those as evidence for global, non-retinotopic feedback. This comparison is especially interesting with respect to scenario one, as a global effect of feedback offers a potential mechanism of few feedback connections exerting a strong influence over V1 processing.

References

Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *Journal of Cognitive Neuroscience*, *15*(4), 600–609.

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., … Rosen, B. R. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(2), 449–454.

Barone, P., Batardiere, A., Knoblauch, K., & Kennedy, H. (2000). Barone, P., Batardiere, A., Knoblauch, K., & Kennedy, H. (2000). Laminar distribution of neurons in extrastriate areas projecting to visual areas V1 and V4 correlates with the hierarchical rank and indicates the operation of a distance rule. Journal of Ne. *Journal of Neuroscience*, *20*(9), 3263–3281.

Bazin, P.-L., Weiss, M., Dinse, J., Schäfer, A., Trampel, R., & Turner, R. (2014). A computational framework for ultra-high resolution cortical segmentation at 7 Tesla. *Neuroimage*, *93*, 201–209.

Benson, N. C., Butt, O. H., Brainard, D. H., & Aguirre, G. K. (2014). Correction of distortion in flattened representations of the cortical surface allows prediction of V1-V3 functional organization from anatomy. *PLoS Computational Biology*, *10*(3).

Brooks, J. C. W., Faull, O. K., Pattinson, K. T. S., & Jenkinson, M. (2013). Physiological noise in brainstem FMRI. *Frontiers in Human Neuroscience*, *7*, 623.

de-Wit, L. H., Kubilius, J., Wagemans, J., & de Beeck, H. P. O. (2012). Bistable Gestalts reduce activity in the whole of V1, not just the retinotopically predicted parts. *Journal of Vision*, *12*(11), 12.

Dumoulin, S. O., & Wandell, B. A. (2008). Population receptive field estimates in human visual cortex. *Neuroimage*, *39*(2), 647–660.

Felleman, D. J., & Van, D. C. E. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex (New York, NY: 1991)*, *1*(1), 1–47.

Goense, J., Merkle, H., & Logothetis, N. K. (2012). High-resolution fMRI reveals laminar differences in neurovascular coupling between positive and negative BOLD responses. *Neuron*, *76*(3), 629–639.

Goffaux, V., Peters, J., Haubrechts, J., Schiltz, C., Jansma, B., & Goebel, R. (2010). From coarse to fine? Spatial and temporal dynamics of cortical face processing. *Cerebral Cortex*, *21*(2), 467–476.

Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., … Halchenko, Y. O. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data*, *3*(1), 1–9.

Harris, K. D., & Mrsic-Flogel, T. D. (2013). Cortical connectivity and sensory coding. *Nature*, *503*(7474), 51–58.

Havlicek, M., & Uludağ, K. (2020). A dynamical model of the laminar BOLD response. *Neuroimage*, *204*, 116209.

Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron*, *36*(5), 791–804.

Huber, L., Ivanov, D., Handwerker, D. A., Marrett, S., Guidi, M., Uludağ, K., … Poser, B. A. (2018). Techniques for blood volume fMRI with VASO: From low-resolution mapping towards sub-millimeter layer-dependent applications. *Neuroimage*, *164*, 131–143.

Hupé, J. M., James, A. C., Payne, B. R., Lomber, S. G., Girard, P., & Bullier, J. (1998). Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. *Nature*, *394*(6695), 784–787.

Juan, C.-H., & Walsh, V. (2003). Feedback to V1: a reverse hierarchy in vision. *Experimental Brain Research*, *150*(2), 259–263.

Kanizsa, G. (1976). Subjective contours. *Scientific American*, *234*(4), 48–53.

Kashyap, S., Ivanov, D., Havlicek, M., Sengupta, S., Poser, B. A., & Uludağ, K. (2018). Resolving laminar activation in human V1 using ultra-high spatial resolution fMRI at 7T. *Scientific Reports*, *8*(1), 1–11.

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annu. Rev. Psychol.*, *55*, 271–304.

Kersten, D., & Yuille, A. (2003). Bayesian models of object perception. *Current Opinion in Neurobiology*, *13*(2), 150–158.

Kobatake, E., & Tanaka, K. (1994). Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *Journal of Neurophysiology*, *71*(3), 856–867.

Kok, P., Bains, L. J., van Mourik, T., Norris, D. G., & de Lange, F. P. (2016). Selective activation of the deep layers of the human primary visual cortex by top-down feedback. *Current Biology*, *26*(3), 371–376.

Koopmans, P. J., Barth, M., Orzada, S., & Norris, D. G. (2011). Multi-echo fMRI of the cortical laminae in humans at 7 T. *Neuroimage*, *56*(3), 1276–1285.

Kubilius, J., Wagemans, J., & Op de Beeck, H. P. (2014). A conceptual

framework of computations in mid-level vision. *Frontiers in Computational Neuroscience*, *8*, 158.

Markov, N. T., Vezoli, J., Chameau, P., Falchier, A., Quilodran, R., Huissoud, C., … Ullman, S. (2014). Anatomy of hierarchy: feedforward and feedback pathways in macaque visual cortex. *Journal of Comparative Neurology*, *522*(1), 225–259.

Marquardt, I., Schneider, M., Gulban, O. F., Ivanov, D., & Uludağ, K. (2018). Cortical depth profiles of luminance contrast responses in human V1 and V2 using 7 T fMRI. *Human Brain Mapping*, *39*(7), 2812–2827.

Muckli, L., De Martino, F., Vizioli, L., Petro, L. S., Smith, F. W., Ugurbil, K., … Yacoub, E. (2015). Contextual feedback to superficial layers of V1. *Current Biology*, *25*(20), 2690–2695.

Op De Beeck, H., & Vogels, R. (2000). Spatial sensitivity of macaque inferior temporal neurons. *Journal of Comparative Neurology*, *426*(4), 505–518.

Polimeni, J. R., Fischl, B., Greve, D. N., & Wald, L. L. (2010). Laminar analysis of 7 T BOLD using an imposed spatial activation pattern in human V1. *Neuroimage*, *52*(4), 1334–1346.

Poort, J., Self, M. W., Van Vugt, B., Malkki, H., & Roelfsema, P. R. (2016). Texture segregation causes early figure enhancement and later ground suppression in areas V1 and V4 of visual cortex. *Cerebral Cortex*, *26*(10), 3964–3976.

Rossion, B., Hanseeuw, B., & Dricot, L. (2012). Defining face perception areas in the human brain: a large-scale factorial fMRI face localizer analysis. *Brain and Cognition*, *79*(2), 138–157.

Scheeringa, R., & Fries, P. (2019). Cortical layers, rhythms and BOLD signals. *NeuroImage*, *197*, 689–698.

Shushruth, S., Ichida, J. M., Levitt, J. B., & Angelucci, A. (2009). Comparison of spatial summation properties of neurons in macaque V1 and V2. *Journal of Neurophysiology*, *102*(4), 2069–2083.

Stojanoski, B., & Cusack, R. (2014a). Time to wave good-bye to phase scrambling: Creating controlled scrambled images using diffeomorphic transformations. *Journal of Vision*, *14*(12), 6.

Stojanoski, B., & Cusack, R. (2014b). Time to wave good-bye to phase scrambling: Creating controlled scrambled images using diffeomorphic transformations. *Journal of Vision*, *14*(12), 6. Retrieved from http://dx.doi.org/10.1167/14.12.6

Torralba, A., & Oliva, A. (2003). Statistics of natural image categories.

Network: Computation in Neural Systems, *14*(3), 391–412.

Tsao, D. Y., Freiwald, W. A., Knutsen, T. A., Mandeville, J. B., & Tootell, R. B. H. (2003). Faces and objects in macaque cerebral cortex. *Nature Neuroscience*, *6*(9), 989–995.

Von der Heydt, R., Peterhans, E., & Baumgartner, G. (1984). Illusory contours and cortical neuron responses. *Science*, *224*(4654), 1260–1262.

Waehnert, M. D., Dinse, J., Weiss, M., Streicher, M. N., Waehnert, P., Geyer, S., … Bazin, P.-L. (2014). Anatomically motivated modeling of cortical laminae. *Neuroimage*, *93*, 210–220.

Wang, L., Mruczek, R. E. B., Arcaro, M. J., & Kastner, S. (2015). Probabilistic maps of visual topography in human cortex. *Cerebral Cortex*, *25*(10), 3911–3931.

Williams, M. A., Baker, C. I., De Beeck, H. P. O., Shim, W. M., Dang, S., Triantafyllou, C., & Kanwisher, N. (2008). Feedback of visual object information to foveal retinotopic cortex. *Nature Neuroscience*, *11*(12), 1439.

Chapter 6
General discussion

Summary

This thesis contains investigations into the processing of visual stimuli in the human brain. The overarching question that guided the research efforts presented here was whether and how the visual system uses feedback to account for statistical regularities in the environment. We used EEG, MEG and fMRI recordings in combination with different image manipulations to address this question. In chapter two we found, in concurrence with previous research, that the basic statistics of visual stimuli shape the way the visual system responds to them, even in the absence of any meaningful semantic information. We showed in chapter three that the presence of low spatial frequencies (LSF) in broadband images of faces reduces the extent to which the scalp distribution of EEG responses resembles that measured during high spatial frequency (HSF) image processing, but not if images were scrambled in phase. We concluded from this reduction in HSF dominance that LSF input might facilitate the efficient processing of HSF content, when the former contains information about the latter. In chapter 4 we followed up on those findings. First, we replicated our previous results in a different group of participants, with slightly different stimulus manipulations and using MEG rather than EEG recordings. We found the same pattern of reduced HSF dominance when informative LSF contrast was available. We then extended our previous findings by using source reconstructions to estimate the influence of LSF information on the neural representations in early visual and higher-level visual cortex respectively. We found reduced gamma band responses in EVC when images of faces contained broadband spatial frequencies rather than HSF alone. This suggests that the presence of LSF information in broadband images alleviates local processing burden. We further found that alpha and beta power, most commonly associated with feedback signaling, in the fusiform gyrus correlates with a decrease in gamma power in EVC for broadband images of faces, but not for broadband phase-scrambled images or images containing only HSF. We consider this as evidence for an LSF driven feedback influence of high-level regions on processing in EVC. Finally, in chapter 5 we investigated the laminar profile of primary visual cortex (V1) responses to a diverse set of object images that were either left intact or distorted to make them unrecognizable and presented to participants at three levels of phase coherence. Preliminary results indicate that V1 differentially responds not only to

varying degrees of phase coherence but also to the presence or absence of semantic information in an image, a property most commonly associated with higher-level visual areas. We argue that the differences in V1 responses between intact and warped images can be attributed to high-level feedback after accounting for differences caused by phase scrambling, which are suitable to disrupt feedback from mid-level visual areas.

The validity of each of the studies presented here first and foremost depends on well controlled image manipulations. In vision research in general, inappropriate use of e.g. Fourier-space filters or ill-adapted stimulus presentation equipment can invalidate entire studies. I will therefore discuss issues with image manipulations as they relate to the studies presented here in the following subsection. This section assumes the reader is familiar with the use of the fine discrete 2D Fourier transform (dFFT2) in digital image processing or has read the introduction to image processing at the end of chapter one of this thesis.
The appropriate use of neuroimaging methods and the interpretation of data derived from indirect measures of brain activity holds further caveats that will be discussed in subsections below.

Stimuli for vision research

Fourier-space filters for spatial frequency and orientation

When producing images for research, it is often desired to restrict spatial frequency and/or orientation profiles to very narrow ranges. 2D Fourier-space filters suffer from similar artifacts as their 1D counterparts and often require tradeoffs between precision in the frequency or orientation response and excessive noise leakage. For example, an ideal, 'brick wall' filter that has sharp cutoffs for both spatial frequency and orientation passbands, would produce ringing artifacts (see figure 1) that might be difficult to see in noise images.
It is therefore desirable to construct filters with smoothly varying (i.e. tapered) edges (for a more detailed introduction to digital image processing see Solomon and Breckon 2011). Figure 2a illustrates the difference between an image produced by a tapered filter and one produced by a brick wall filter.

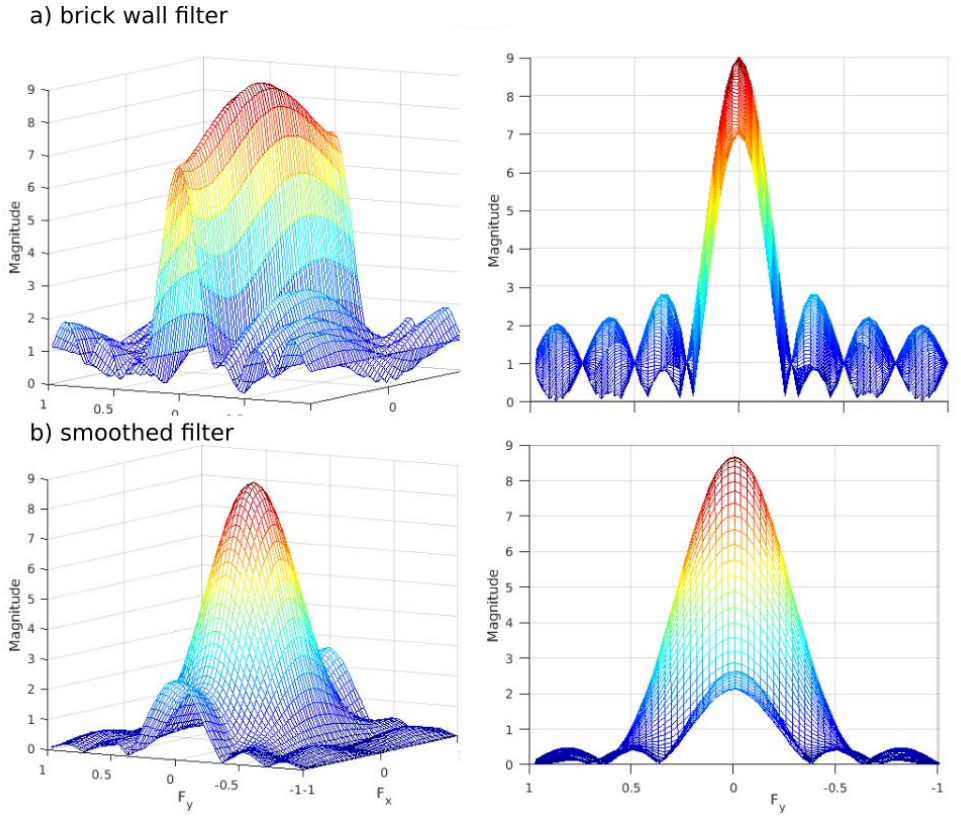


Figure 1: Filtering images in orientation, spatial frequency, or both, requires a tradeoff between filter precision and filter artifacts. High precision in e.g. the frequency-passband requires steep transitions between passed and attenuated frequencies. However, such steep transitions lead to ripples in the filters impulse response (upper panel) that occur because the filter output oscillates between higher and lower values than the maximum and minimum of the filter input (Ogata 1995). *A smooth transition band is less frequency selective (see the larger spread FWHM in (b) compared to (a)), but also produces fewer and lower amplitude ripples.*

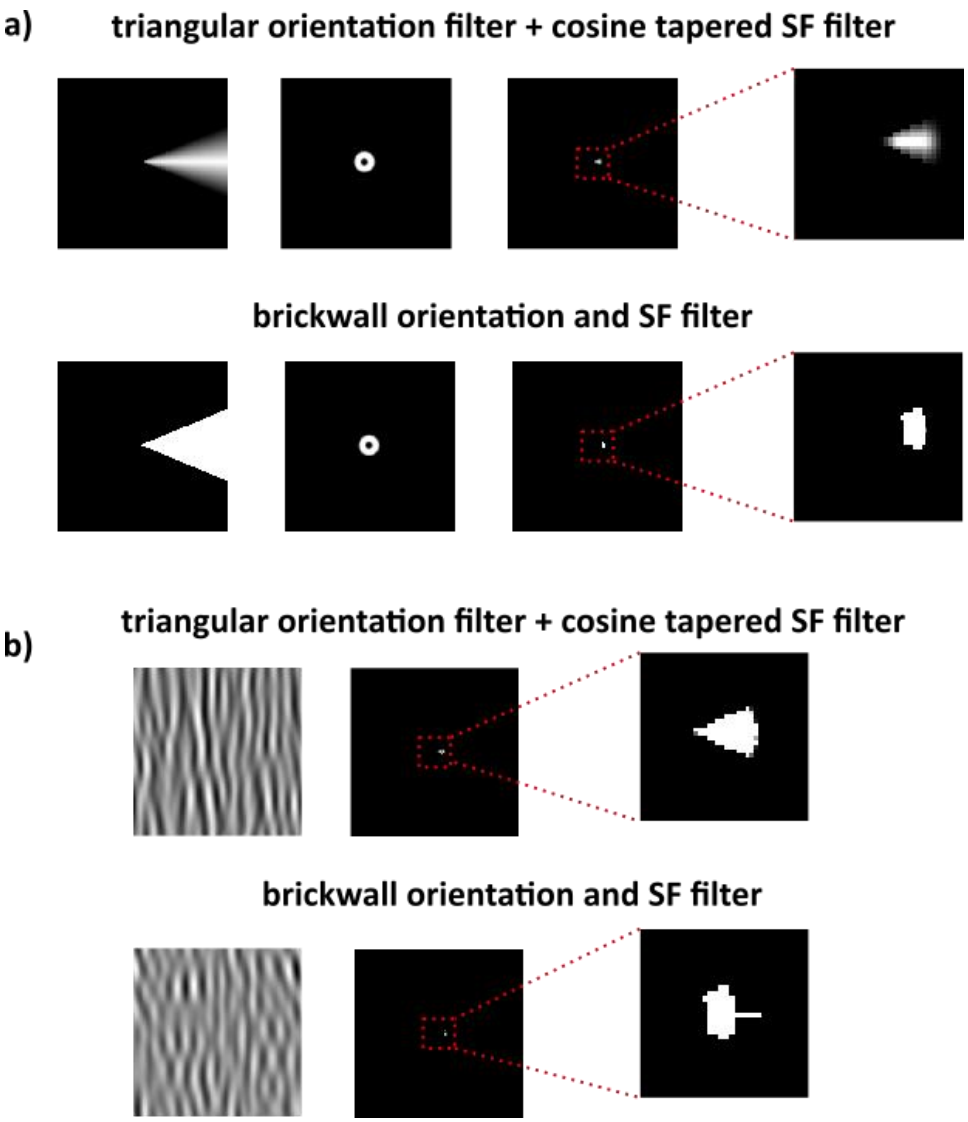


Figure2: Brick wall filters produce excessive amounts of image artifacts and are rarely suitable for producing stimuli for vision science. a) shows combined orientation and spatial frequency filters produced via a triangular orientation filter, were the passband is tapered-off linearly from the center orientation and a cosine tapered annulus SF filter (upper row) and the same filters without any tapering (lower row). b) shows the resulting image after multiplying the filters with the same white noise image in the Fourier domain as well as the spectrum those images produce when applying dFFT2 to them.

The choice of filter-type determines the shape of the tapered filter-edge and therewith the frequency and orientation profile of the filtered image. The commonly used gaussian filter for example produces an impulse response that follows the characteristic bell shape with no sharp edges and therefore produces no ringing artifacts if properly scaled. However, because the filter has no plateau, frequencies are passed at different magnitudes, with a drop-off from the center of the filter kernel. For the studies in chapters three and four, we required the spatial frequencies in the passband to be attenuated as little as possible. We hence used 7th order Butterworth filters, that are designed to have a maximally flat passband, while still avoiding sharp edges. However, even well tapered filters might not produce the expected result with respect to the orientation spectrum. This is because the dFFT2 does not sample all orientations equally. A continuous FFT generates an amplitude spectrum with a vector for each possible orientation in polar coordinates (i.e. after shifting to center on the zero-frequency). Each point on those vectors corresponds to the amplitude of the vector's orientation at a specific spatial frequency. Figure 3 demonstrates that on a finite pixel grid, as produced by the discrete FFT, orientations that are not parallel to the cardinal axes (0° or 90°) and not 45° or 135° oblique, are undersampled, especially at lower frequencies.

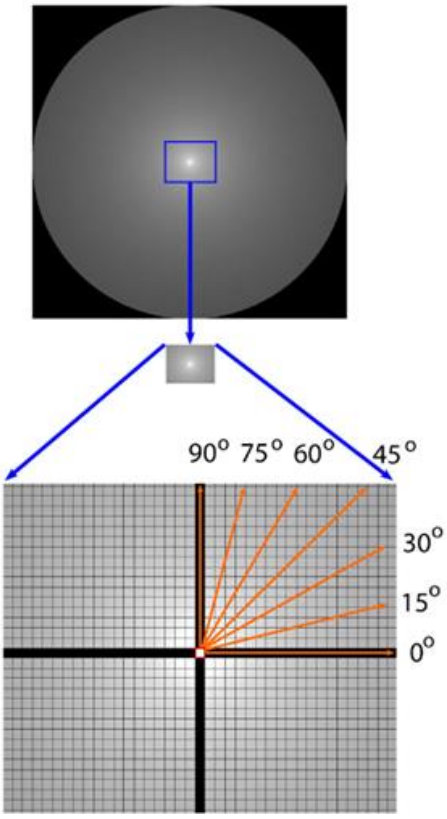


Figure 3: Schematic representation of the lack of a set of continuous orientation vectors in the Fourier domain (i.e., the discrete Fourier domain amplitude spectrum). Top: Example of an isotropic amplitude spectrum with a region of interest highlighted in blue. Middle: The same region of interest shown in isolation. Bottom: The same isolated region of interest enlarged with a matrix grid laid on top. This grid represents the nature of the discrete sampling incurred by the discrete Fourier transform. All of the 0° and 90° coordinates have been shaded in black for clarity. The center square (white, highlighted in red) represents the DC component. The orange arrows represent orientation vectors drawn at arbitrary orientations in steps of 15°. Notice how the 0°, 45°, and 90° vectors contain sampling points (i.e., centers of the squares) along their entire lengths, whereas the other vectors do not. Reprinted with caption from Hansen & Essock (2004)

Spectral limits of digital images

In chapter two, we utilized stimuli that were restricted to spatial frequency bandwidths of only one octave and orientation bandwidths of 21°. To avoid potential inaccuracies due to filter cutoffs or orientation misrepresentation, we produced our stimuli from individual sine wave gratings as described in Haun & Essock (2010). While producing the corresponding single sine gratings is comparatively straightforward, the narrow band gratings require careful calibration. The combination of many sinewave gratings of different phase, SF and orientation will almost inevitably lead to most of those gratings containing non-integer numbers of cycles within the image, which in turn leads to inaccuracies when trying to estimate the power at those frequencies. Furthermore, overlaying sine wave gratings with different phase angles and orientations produces regions where peaks fall on peaks and troughs fall on troughs. Not only does this result in the introduction of new spatial frequencies and orientations, but it produces regions that exceed the maximum representable luminance. Re-scaling the image luminance to the representable range of 0-1 however, reduces the contrast of the image as with only a limited number of luminance steps available, image detail is unavoidably lost. In our study, we mitigated the loss of image information by using the grayscale extension available with a V-Pixx projector (VPixx Technologies, Canada) in M16 mode, enabling 10-12 bits of grayscale resolution. In M16 mode, a 16-bit shade of gray is calculated and transmitted to the projector using the 8 bits of red and 8 bits of green in an RGB image. The device then combines those to display a 10-12 bit closest value of gray. Nevertheless, achievable contrasts are lower for narrow band composite sine gratings than for single sine gratings. We therefore had to limit the dynamic range of single sine gratings to match those of the narrow-band images. Furthermore, we were only able to use the full extent of the dynamic range for narrow-band images in high overall luminance settings. It has been shown that human spatial frequency sensitivity shifts towards higher frequencies in high luminance environments (Campbell and Robson 1968). Therefore, it is possible that the high mean luminance of our stimuli (300cd/m^2) resulted in sensitivity to higher spatial frequencies than the more commonly used intermediate mean luminance values around 90-150cd/m^2 (Bühren et al. 2006; Cox, Norman, and Norman 1999).

Contrast and spatial frequency

In chapters three and four we used stimuli that were cropped to an oval shape and then manipulated to contain only low, only high or a combination of low and high spatial frequencies. Previous studies investigating spatial frequency dependent processing of complex stimuli frequently equated image contrast across SF ranges

to account for the otherwise higher contrast of LSF (Vlamings, Goffaux, and Kemner 2009; Peyrin et al. 2006; Valérie Goffaux and Rossion 2006; Mu and Li 2013). However, this deviation from natural, $1/SF^{\alpha}$ amplitude spectral properties might bias visual perception. Kauffmann et al. (2015) showed that both behavioral and fMRI responses to natural scenes presented with different spatial frequency differed dependent on whether scene images were equalized for root mean square (RMS) contrast or not. In order to avoid differences in image contrast between SF conditions, while preserving a naturalistic slope of the amplitude spectrum, we chose the cutoffs of our frequency filters such that the area under the curve of the amplitude spectrum was equal to either side of the cutoff. To prevent the LSF images from being dominated by very low frequency contrast with little informational value, we excluded SFs up to 5 cycles per image (cpi). While this manipulation equates RMS contrast across stimulus conditions and preserves the $1/ SF^{\alpha}$ characteristic of the amplitude spectrum, it also leads to the loss of the very low spatial information that is thought to contribute most to the coarse-to-fine (CtF) processing we wanted to investigate (e.g. in Valerie Goffaux et al. 2010 LSF reached from 2 to 8 cpi), as well as vastly different bandwidths between conditions, with lower bandwidths for LSF images. In our studies, the LSF passband only reached from 5cpi to 8cpi whereas the HSF passband included all frequencies between 8cpi and the Nyquist frequency. The loss of very low SF is not a mayor concern in our study design, as a reduction of CtF would not benefit our hypothesis and is therefore unlikely to cause any false positives. Accepting large differences in bandwidths however was certainly a compromise, because it could have led to a reduction of image complexity in LSF conditions. In relation to the overarching theme of this thesis, namely whether and how the visual system exploits redundancies in visual input to aide visual perception, differences in image complexity could undermine the validity of our study. The issue is therefore discussed in more detail below.

Image complexity

Although image analyses showed that image entropies, as a simple proxy of image information, are virtually identical across SF conditions in the stimuli used in chapters three and four, other, more ecologically relevant measures of image complexity do reveal differences between conditions. In the Olshausen and Field (O&F) model (Olshausen and Field 1997), where images are assumed to be composed of a linear superposition of basis-functions that are mixed with given amplitudes, the number of basis-functions required to account for 90% of the image variance differs between LSF and HSF images (Matlab code for model generation is available for download at http://www.rctn.org/bruno/sparsenet/). The O&F model is based on the notion that primary visual cortex receptive fields

with their tuning to spatial frequency and orientation decompose information similar to the basis-functions derived from wavelet transforms and with those could produce a sparse code of image representations. In such a distributed sparse coding regime, sensory information is represented by a small number of neurons at a time which are highly dispersed across space (see Olshausen and Field 2004 for review). Empirical evidence supports the notion of a distributed sparse representation of natural images in V1. Vinje & Gallant (2000) recorded from V1 neurons of macaque monkeys while they were freely viewing natural scenes. They found that responses between pairs of neurons were strongly decorrelated with stimulation of their non-classical receptive fields (i.e. the inhibitory surround of an excitatory receptive field). Such stimulation further increased selectivity of individual V1 neurons as well as the sparseness of the population response distribution. In such a regime, LSF images, that can be represented by fewer basis-functions, would require fewer resources than HSF images, i.e. the images themselves could be coded more sparsely. Comparing neural responses based on the combined activity of large populations of cells, as measured with EEG or MEG, would therefore suffer from sparseness confounds. In our studies, we do not compare responses to LSF and HSF images directly, but rather investigate how higher-level image structure (i.e. the presence of a face) influences the relative dominance of LSF and HSF representations by using Fourier phase-scrambled images with the same SF and orientation content as the original images for comparison. While restricting interpretation of relative SF dominance to the difference between intact and scrambled image processing avoids potential confounds caused by unequal image complexity between SF ranges, phase scrambling itself could also cause differences in complexity.

Fourier phase scrambling modulates primary stages of visual processing

Basic models of the primary visual cortex, such as the HMAX model of object perception (Riesenhuber and Poggio 2002) that relies solely on feedforward projections, make no difference between intact and phase-scrambled images[1] at model stages representative of primary visual cortex. Image entropy as well as the number of basis-functions required to reconstruct both images are identical. Yet, studies in humans and animals found pronounced response disparities between the

[1] *In Stojanoski & Cusack* (2014) *the authors do find a significant difference between intact and phase scrambled images in layer C1 of the HMAXX model. However, this is likely due to the fact that the objects used in that study are segmented images on white background and their phase scrambled versions are not restricted to the same area as is covered by the intact object. As we show in chapter 3, the dFFT2 takes the entirety of the image space into account. The phase scrambled version of a segmented object is therefore contaminated by its uniform background which introduces spurious low spatial frequencies. This problem can easily be remedied by iteratively scrambling and re-pasting the intact figure until SF differences between intact and scrambled images fade.*

two conditions (Arsenault, Yoonessi, and Baker 2011; Rainer et al. 2002; Coggan et al. 2017; Murray et al. 2002; Valerie Goffaux et al. 2016) in perceptual ratings and fMRI studies of early visual cortex. However, the fMRI studies were conducted with conventional resolutions (voxel size >1mm^3), where feedforward and feedback influences on V1 processing are confounded (see section 6.4). Image phase coherence may directly affect the processing of feedforward input in V1. Alternatively, the higher-level processing stages that are sensitive to the redundancies in intact image content may produce feedback signals, which streamline V1 processing. This notion is supported by Murray et al. (2002). There, the authors found reduced V1 fMRI BOLD in response to objects compared to a random arrangement of the object's elements. They further found the reduction of V1 activation to be paralleled by an increase in activity in the lateral occipital complex (LOC) for intact objects. They interpret their findings as reflecting a feedback influence of higher-level visual cortex on V1.

The study presented in chapter five more explicitly considers Fourier phase coherence as potentially influencing perception at processing stages prior to high-level category-selective cortex. We used 7T fMRI to sample primary visual cortex at different levels of cortical depth to dissociate feedforward from feedback effects. We applied two different manipulations to obscure image content: Fourier phase scrambling and diffeomorphic transformations. Phase scrambling, where the Fourier phase information is randomized, removes all but the most basic image features and is therefore thought to disrupt processing at mid-level visual regions (i.e. V2-V4) (Freeman et al. 2013). Diffeomorphic transformations, where the image is distorted by expanding and contracting pixels based on two dimensional cosine components, each with random amplitude and phase, are thought to preserve image properties mid-level regions are selective to (Stojanoski and Cusack 2014). Those stimuli have been found to produce similar activations as intact images up to mid-level visual regions within the posterior regions of the ventral pathway (Freud et al. 2017). The authors presented participants with images that were progressively more distorted as well as images that were progressively more box-scrambled (i.e., image is subdivided into blocks that are subsequently randomized in their order) and compared voxel responses as a function of object coherence. They found the voxel activation along the visual pathway to increase with increasing image coherence. Importantly, the slope of this increase was steeper for box scrambled than for diffeomorphic images up to area LOC, which has previously been found to respond to the general presence of coherent shapes and objects, but not to their semantic content (Malach et al. 1995; Margalit et al. 2016). Accordingly, the assumption of our study is that phase scrambling disrupts processing in both low- to mid-level and high-level category-selective regions, whereas diffeomorphic transformations selectively disrupt processing in the latter.

Because operations in Fourier space affect all pixels of the image in cartesian space, some additional care is required to avoid spectral mismatches between intact and scrambled objects when they are cropped to a non-square contour. In the image processing of chapters 3 to 5, we aimed to reduce potential confounds due to the amount of space in pixels occupied by stimuli of all conditions, as well as their spatial frequency and phase spectra. In order to ensure that our phase-scrambled objects occupy the same space in the image and contain the same amplitude and orientation spectrum, we produced masks for all images, which we used to create image regions of interest (iROIs) covering the object and excluding the background. For this, we produced images of grayscale objects as RGB matrices and made the background green. The green background then allows to unambiguously tag pixels as belonging to the figure or to the background. We then produced Fourier phase-scrambled versions of each object by iteratively scrambling the whole image and re-pasting the object as indexed by the pre-calculated iROIs. This procedure removes SF differences between intact and scrambled objects that are found after a single iteration of the scrambling procedure due to the uniform (i.e. LSF) background.

When generating intermediate levels of phase coherence, as we did in chapter 5, it is further important to take the circular nature of the distribution of phase into account. Creating progressively scrambled images via linear interpolation between images with intact phase and images with uniform random phase produces distributions that over-represent certain phases and have non-uniform step sizes between phase angles (Dakin et al. 2002; Sadr and Sinha 2004). To avoid those issues and preserve a uniform distribution of phase, we linearly interpolated phase angles towards the direction of the smallest distance between angles as suggested by Ales et al. (2012).
Although the so produced images were well matched for luminance, contrast and SF spectrum, they differed substantially in properties such as contour length, curvature (the diffeomorphic manipulation tends to produce rounded contours) and entropy, which might have had unforeseen consequences for early or mid-level visual processing.

In summary, although we, as much as any other vision scientist, tried our best to produce well controlled stimuli for our experiments, image manipulations with the methods currently available often require tradeoffs. Therefore, the conclusions drawn from any individual experiment, including the ones presented in this thesis, should be considered with caution, and compared to similar experiments using different manipulations.
The same concerns hold for other areas of methodology as discussed below.

Measurement and information

In chapter 3 we applied support vector machine (SVM) classification to the time varying patterns of EEG scalp distributions. We trained classifiers to assign condition labels to scalp distribution patterns over occipital channels dependent on trials belonging to either the LSF noise or the HSF noise condition. We then investigated the classifiers' generalization performance both across time and across stimulus condition. We interpreted the response a LSF/HSF trained classifier gave on broadband trials were LSF and HSF contributed equally to image contrast, as reflecting a representational dominance of the chosen class. There are technical as well as theoretical concerns to consider in this interpretation. The technical concern is about the way an algorithmic classifier, such as the SVM used in this study or the linear discriminant analysis (LDA) used in the study shown in chapter 4, arrives at a classification result. The theoretical concern is more generally related to how measured brain activity links to the representation of information in the brain. The following will discuss both issues.

How does a classifier classify?

Multivariate pattern analysis has been introduced into the neuroscientific literature in the early 2000s (Haxby et al. 2001) as a way to overcome the limitations mass univariate contrasting introduces, namely the discarding of weaker responses as well as their distribution when averaging over many voxels in fMRI analysis (see Haxby 2012 for the authors own account). Today, multivariate analysis, often used in the form of pattern classification, has become a popular tool in fMRI, EEG and MEG analysis (see figure 4). Despite the many easy-to-implement software packages available to perform classification analysis (Bode et al. 2019; http://cosmomvpa.org; https://github.com/treder/MVPA-Light etc.) optimization procedures used to enhance classification results, be it the form of feature selection or in the form of choosing regularization parameters to prevent overfitting the data may introduce spurious findings in classification analyses (Varoquaux et al. 2017) and need to be chosen with care. The linear *SVM* algorithm we used in chapter three produces a split of two classes in high dimensional space by maximizing the margin around the hyperplane separating them (Cortes and Vapnik 1995). Above chance classification performance indicates that, at least to some extent, the data are linearly separable into two classes, meaning that at least one feature reliably differs between conditions. However, there is not one unique combination of feature differences that gives rise to one particular level of classification accuracy i.e. any number of feature combinations could result in the same overall accuracy.

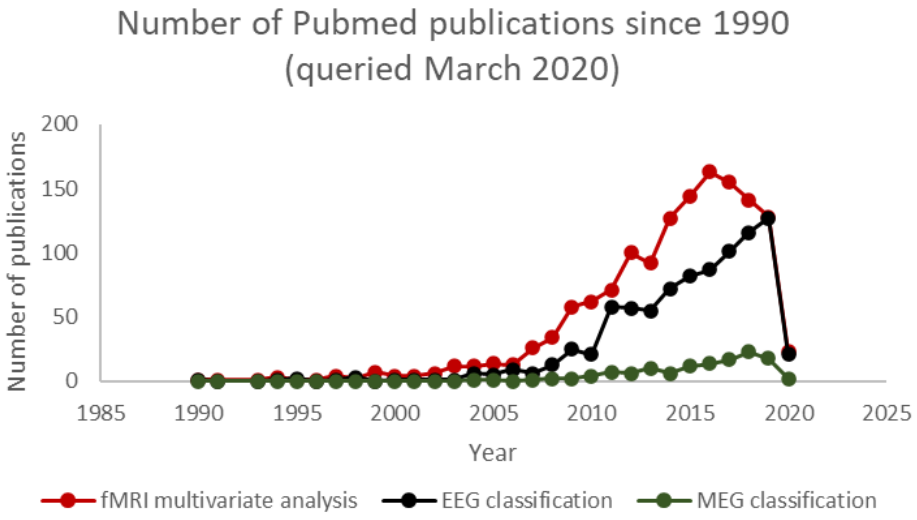


Figure 4: Number of pubmed listed publication with keywords fMRI + multivariate analysis, EEG + classification and MEG + classification since 1990 (accessed on March 1st, 2020).

There have been attempts to map feature weights, i.e. how much a given feature contributed to the classifiers' decision, in order to gain information about the cortical sources driving class separability (Bode et al. 2019). However, those weights might be affected by statistically independent signals that carry a high weight, but are irrelevant to the desired class separation, such as high amplitude muscle-noise. The algorithm developed by Haufe et al. (2014) and implemented in Bode et al. (2019) can in some cases mitigate this misinterpretation of feature weights by projecting weight vectors onto activation maps prior to interpretation. In most neuroimaging studies however, weight interpretations remain problematic even after vector projection onto activation maps as demonstrated by Douglas and Anderson (2019). They show that due to noise in the data, interactions between features, as well as information redundancies introduced by feature selection processes, weights might still favor noise features over informative ones.
Independent of the exact classifier specifications, a 2015 paper raised concerns about the statistical evaluation of classification results (Combrisson and Jerbi 2015). The authors state that in neuroimaging classification, where sample sizes tend to be small, statistical evaluation of classification results against *a priori* chance levels (e.g. 50% and 25% for 2 or 4 class classification respectively) leads to false positive results. The paper has caused animated debates and the authors' claim that evaluating classification results based on *a priori* chance levels without proper permutation procedures is common practice in neuroscientific publications has

since been refuted (see linked discussion in footnote[2]). Nevertheless, the issue of how to perform statistical testing on classification results is worth considering. In our studies, we evaluate the statistical significance of above chance classification performance against empirical chance. For this, we generate data with permuted condition labels within all cross-validation steps before repeating cross-validation as suggested by Hebart in the discussion linked in footnote [2].

In our studies we further used SVM and LDA in combination with the temporal generalization method, introduced by King and Dehaene (2014). Temporal generalization of classifiers produces a training-time by testing time matrix where the diagonal reflects training data and testing data from different trials being taken from corresponding samples of trial time with respect to some event (e.g. 100ms after stimulus onset). Matrix positions away from the diagonal correspond to training and testing data coming not only from different trials, but also different time samples within those trials. In their original paper the authors applied their method to a range of simulated cortical generators with different temporal response profiles to demonstrate that the patterns of temporal generalization performance are suited to capture the processes underlying scalp projections of generator activity. But importantly, the authors do not show temporal generalization patterns to *uniquely* reflect a specific temporal dynamic, i.e. while a specific pattern of generator activity can be linked to only one pattern of temporal generalization matrix, the reverse is not certain. Ergo, drawing strong conclusions about the dynamics of cortical signal generators solely from the temporal generalization matrices is untenable. In our studies, we therefore limit our interpretation of temporal generalization performance as reflecting a classifiers ability to predict class labels across trial-time.

In summary, while the multivariate classification of patterns of neuroimaging data often provides a more sensitive measure for condition differences than the comparison of averages after mass univariate analyses, classification results need to be interpreted with caution.

Information for whom?

A more fundamental problem than finding the correct statistical approach to evaluating pattern classification results is that the patterns of measured activation the classifier uses to achieve a given accuracy in assigning the correct labels to conditions, might not mean anything to the rest of the brain at all.

---

[2] The original discussion in response to a commentary by science blogger Neurosceptic [https://www.discovermagazine.com/blog/neuroskeptic] can be retrieved from the internet archive under https://web.archive.org/web/20151026233534/http://blogs.discovermagazine.com/neuroskeptic/2015/01/18/machine-learning-exceeding-chance-by-chance#more-5985.
See http://mvpa.blogspot.com/2015/01/research-blogging-exceeding-chance.html for refutation of the claims on improper statistical evaluations.

De Wit et al. (2016) argue that for the patterns of activation used as classifier features, or any other measure of brain activity for that matter, to be interpreted as information, the target receiver of that information has to be considered. For example, hypothetical patterns of primary visual cortex activation in response to two conditions measured with a neuroimaging technique might differ in some way the classifier is able to pick up and use to provide an accurate prediction of condition label. So, to the classifier and through it to the experimenter using the classifier, V1 responses contain information about condition class membership. But assume that V2, that receives most of its feedforward input from V1, does not respond to either one of the conditions. From the perspective of V2, V1 does not convey information. Thus, the implicit assumption that decodable activation patterns recorded from the brain represent information only holds if the experimenter is the receiver. The more interesting question however is whether and how the brain can interpret such information. In the framework of Shannon's theory of information (see introduction) where a signal generated by a transmitter is conveyed to a receiver via a noisy channel, it is not sufficient to consider only the signal itself to understand the message within (Shannon 1948). Instead, the properties of the receiver and its ability to extract information from the channel are defining which parts, if any, of the signal can be considered information. In cognitive neuroscience research, this means that interpreting the neural *correlates* of a cognitive process, say object perception, that we can measure in numerous ways, as a neural *representation* of the process, requires additional steps. The study in chapter 3 does not take those additional steps. Although we discuss potential mechanisms of how the visual system could utilize LSF image content to reduce HSF related processing load, we empirically only show that the presence of LSF in BB images *correlates* with a measured pattern of scalp activation that is less similar to a pattern measured during HSF image viewing when images are intact rather than scrambled.

For activity in a given brain region to be considered information, it needs to influence activity in a target (receiver) region. In the study shown in chapter four we extend our previous findings to consider information transfer between regions. We used MEG source reconstruction to form cortical regions of interest and subsequently investigated how activity in regions we considered likely sources of stimulus content driven feedback to early visual cortex (EVC) relates to the activity reconstructed to early visual cortex. In this framework, the feedback sources are the transmitter, the alpha and beta band oscillatory activity we took as a proxy of feedback processes forms the channel and EVC is the receiver. A change in V1 activity related to a preceding change in feedback sources would then allow us to interpret the activity in feedback source regions as representing information (notwithstanding the many other technical and theoretical constraints that apply to such connectivity measures, see paragraph below).

We however still cannot make any strong inferences about the information in EVC itself. To infer on EVC information representation, we need to consider the likely receiver of EVC output. We did this in the study described in chapter 5. There, we measured BOLD responses related to intact and warped images at different levels of image phase coherence. We considered two regions of interest: primary visual cortex (V1) and secondary visual cortex (V2). The high resolution afforded by fMR imaging at ultra-high field (7T) allowed us to separate feedforward and feedback information channels (but see discussion of cortical depth dependent imaging below for qualifying remarks on the method) between V1 and V2. Thus, changes in V1 response between conditions could be interpreted as representing information if a corresponding change were found in layers related to feedforward activity in V2.

However, even with (indirect) access to the transmitter as well as the receiver of information, what we consider as the channel still depends on our preconceptions and theories. Consider for example the retinotopic maps in EVC, where neighboring regions of the retina are mapped onto neighboring patches of cortex. Given the two shapes in figure 5, fMRI-based pattern classification with sufficiently high resolution could successfully differentiate between the two figures in both V1 and V2. The classifier's decision would be based largely on the difference in voxels covered in the two conditions.

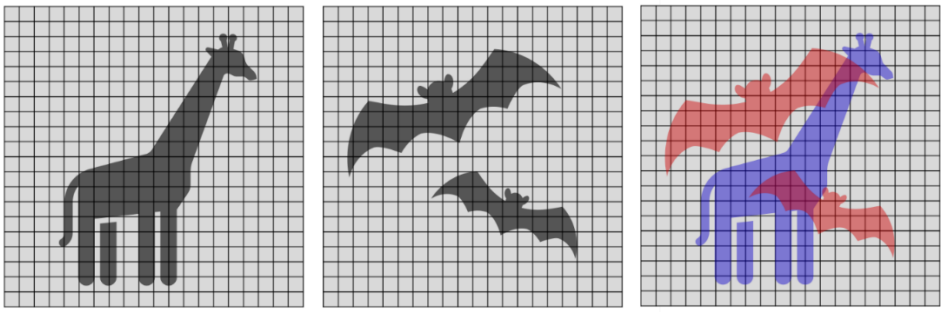


Figure 5: The shapes of the giraffe and the bats differ, among other things, in the voxels they cover on a hypothetical retinotopic map schematized by the grid. If we attempted to classify whether a participant saw the giraffe or the bats on a given trial given the BOLD responses of voxels from retinotopic cortex, the classifiers response would be largely based on the images difference with respect to the voxels covered.

Can the conclusion therefore be that the retinotopic organization of the stimulus in V1 (the transmitter), that is also observed in V2 (the receiver) forms the channel for the information transfer between the two. In other words, if retinotopic

organization is informative to our classifier, must it then also be informative to V2? According to Koenderink (1990) it might not. Instead, he proposes that retinotopic organization in cortex is merely a byproduct of development and information about spatial relationships between stimulus elements is instead contained in correlations between receptive fields within the same retinotopic map. Williams et al. (2007) further demonstrated, that performance of a classifier given V1 responses, presumably largely based on retinotopic organization, does not predict participants' task performance on the same classification task. They presented participants with three classes of visual objects while scanning fMRI at a spatial resolution of 1.4 * 1.4 * 2.0 mm + 20% spacing. Participants were asked to indicate which class each object belonged to. Subsequently, correlations between cortical response patterns to each stimulus class were compared independently for cortical regions including V1 and lateral occipital cortex (LOC). Only misclassification in LOC correlated with behavioral misclassification. Ergo, the patterns of V1 activation used by the classifier and commonly interpreted as representing information, were not representative of the information used by the brain to derive its conclusion about stimulus class. Thus, paradigms that do not consider the (intermediate or final) target of information are an insufficient basis to interpret neuronal activity in terms of information representation in the brain. Instead, Decharms & Zador (2000) suggest, that representation must be characterized by both content and function.

Source reconstruction of EEG and MEG data

Electroencephalography (EEG) and magnetoencephalography (MEG) both provide measures based on electrical currents in the brain. While EEG measures variations in the electric field formed by the postsynaptic flow of current within populations of synchronously activated pyramidal neurons, MEG measures the magnetic field variations that accompany the electric fields (see figure 6 for schematic representation). Both methods can sample electrical activity at very high temporal resolutions, but originally provide poor spatial resolution.

To interpret the cortical sources of the electrical response measures from outside the head, source reconstruction methods have been developed (see Grech et al. 2008 for review of EEG source reconstruction methods as well as Hincapié et al. 2017; and Sekihara, Sahani, and Nagarajan 2005 for MEG).

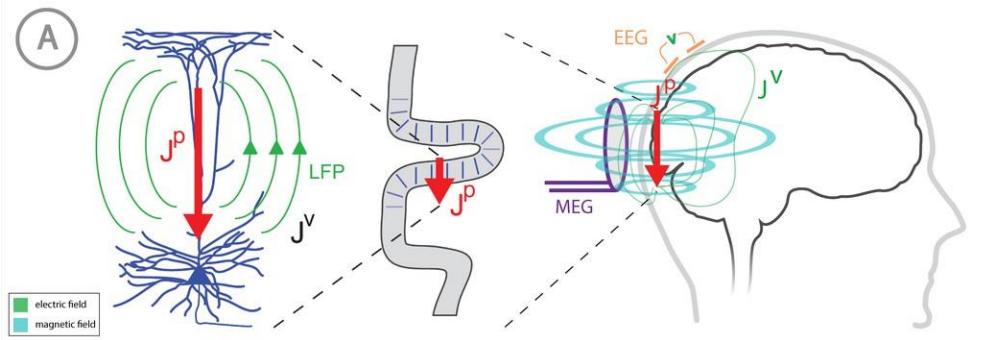


Figure 6: Both EEG and MEG signals are generated by the postsynaptic, intracellular current flow in the long and spatially aligned dendrites of a large population of synchronously activated pyramidal neurons (left). The primary electrical currents (JP, illustrated with red arrow) flowing from the dendrites toward the cell body, give rise to an electric field with field-lines parallel to the direction of current flow which can be picked up by EEG electrodes. At the same time a magnetic field that is orthogonal to the electric field is formed and can be picked up by MEG sensors (Figure adapted from (Neymotin et al. 2020)

EGG source reconstruction, more so than MEG based reconstructions, suffers from signal distortions due to volume conduction (Wolters et al. 2004) and errors in electrode positions (Dalal et al. 2014). Both of our studies containing source reconstruction results (chapters 2 and 4) are therefore based on MEG data and the following will discuss issues with source reconstruction as they relate to MEG.

The reconstruction of the cortical distribution of activity is complicated by the mismatch between the limited number of channels, most of which are also highly correlated, and the large number of possible source locations or distributions of active sources (Hauk, Wakeman, and Henson 2011). To constrain the solution space, assumptions about the sources must be made. One straightforward assumption we make in chapters 2 and 4 is that neuronal activity originates in the brain. Another assumption is that the spread of electrical activity through the brain and the head is determined by the anatomical organization and conductivity of different tissue types. In both studies we produce volume conduction models based on the anatomical MRI of each participant combined with *a priori* conductivity values. We then calculate leadfields, i.e. the predicted magnetic field vector that would be measured outside the scalp if a hypothetical dipole inside the brain was active, for a set of dipole locations. In chapter 4 those dipoles are initially equidistant points on a regular 3-dimensional grid, that is subsequently warped to the shape of the brain and restricted to positions inside the brain. In chapter two, source positions are considered along the vertices of a surface rendering of the midpoint between the gray matter / white matter boundary and the gray matter /

cerebro-spinal fluid boundary and therefore restricted to gray matter. This constitutes the forward model. We then use different approaches, namely linearly constrained minimum variance (LCMV) beamforming and minimum norm estimates (MNE) to invert the forward model, given the data measured during MEG recording. The goal of inverse modeling is to estimate the strength of the current sources that form the measured combination of magnetic field vectors. However, because some of the sources do not produce a measurable change in field vectors and those sources can be added to any model without effecting the data fit (Helmholtz 1853), an infinite number of inverse solutions with equal validity are possible. Further, the forward models used in chapters two and four contained ~15 000 and ~40 000 potential sources respectively, but we only have measured data from 204 highly correlated sensors, thus, the inversion problem is severely underspecified. To mitigate this problem further, assumptions about how the measured data relates to the sources need to be made. MNE used in chapter two assumes that out of all the possible combinations of sources able to explain the measured data, the optimal configuration is the one with the smallest overall source power (Hauk 2004). Importantly, this assumption leads to largely superficial sources because for deep sources to effect sensors in a similar way, high amplitudes would have to be attributed to them which would violate the smallest source power assumption. Appropriate depth weighing needs to be applied to prevent such bias towards superficial sources (Lin et al. 2006).

The LCMV beamformer (Van Veen et al. 1997) used in chapter four operates under the additional assumption that sources that are spatially separate are not strongly correlated (Hillebrand and Barnes 2003), which has been demonstrated to be valid in most situations (Hillebrand and Barnes 2005). This assumption is violated when the stimulation paradigm produces zero, or near-zero lag sources that are symmetric across hemispheres. The long duration (>2seconds) periodic visual stimulation like we used in chapter two and the localizer in chapter four produces such zero-lag correlated sources.

The correlated sources produced by the localizer in chapter four can be recovered by modelling symmetric dipoles and subsequently separating them into a dipole of interest and a suppression dipole (Popov, Oostenveld, and Schoffelen 2018). This is because in our localizer paradigm, participants were presented with flickering images of faces, a stimulus known to evoke a lateralized response (Rossion et al. 2003; e.g. Meng et al. 2012).

The bilateral activations evoked by the flickering simple patterns in chapter two however, do not produce coherent sources over visual areas with the same approach. They are therefore better addressed with MNE, even though substituting beamformers with MNE comes at the cost of less focal sources.

In conclusion, the success of MEG source reconstruction relies on the use of well-informed assumptions to constrain an underspecified problem. If these constrains

are considered, spatially distributed sources can be reliably reconstructed with high temporal and reasonable spatial precision.

Cortical depth dependent fMRI

As a final note to methodological limitations the following will give a brief overview of the current methodological challenges related to the analysis and interpretation of our ultra-high field (UHF) fMRI study.
In chapter five, we use fMR imaging at UHF (7 Tesla) to investigate feedforward and feedback processing in early visual cortex. In animal tracing and electrophysiology studies, the different cortical laminae haven been found to be selectively targeted by neurons from hierarchically lower or higher areas, respectively (Barone et al. 2000; Markov et al. 2014). Feedforward (FF) input originates in layer 3 and largely targets granular layer 4. Feedback (FB) input from proximal and distal areas targets supragranular layers 2/3 and infragranular layer 6 respectively (see figure 7).

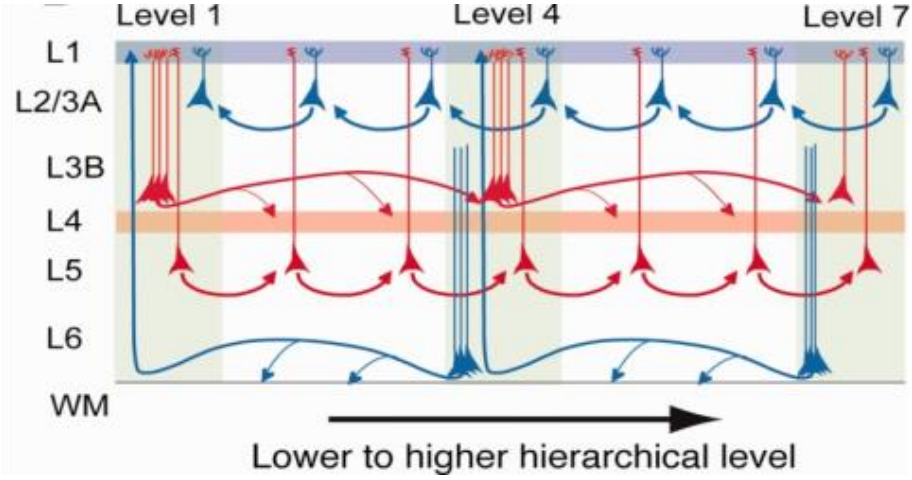


Figure 7: Schematic representation of FF and FB signaling between hypothetical cortical areas Level 1 to Level 7. Red represents FF and blue represents FB. FF neurons have their cell bodies in layer 3 (and some in layer 5), their dendrites in layer 1 and their axons project to granular (and some to layer 5). FB neurons project from and to supragranular layers 2/3A for proximal and from and to infragranular layer 6 for distal connections. Figure is reprinted from Markov et al. (2014)

UHF imaging holds the promise of making those feedforward and feedback dominated cortical layers accessible for investigation on a large scale and in a non-invasive manner. But as usual, it's not that simple.

Because the gyromagnetic ratio, i.e. the ratio of magnetic moment to angular momentum, is a constant for any given nucleus (e.g. 42.58 MHz/T for hydrogen nuclei), the necessary pulse frequency to excite a nucleus can be precisely calculated given the strength of the static magnetic field (B_0) via the Larmor equation ($\omega = \gamma B_0$ where ω is the frequency in mHZ, γ is the gyromagnetic ratio in mHz/Tesla and B_0 is the magnetic field strength in Tesla). By manipulating the B_0 e.g. by introducing a linear decrease in field strength along one dimension of the brain, the effect of a given RF pulse can be limited to arbitrarily thin slices of brain tissue. As the hydrogen nuclei excited by the RF pulse return to their preferred low energy state, they emit a signal that is picked up by receiver coils (see figure 8).

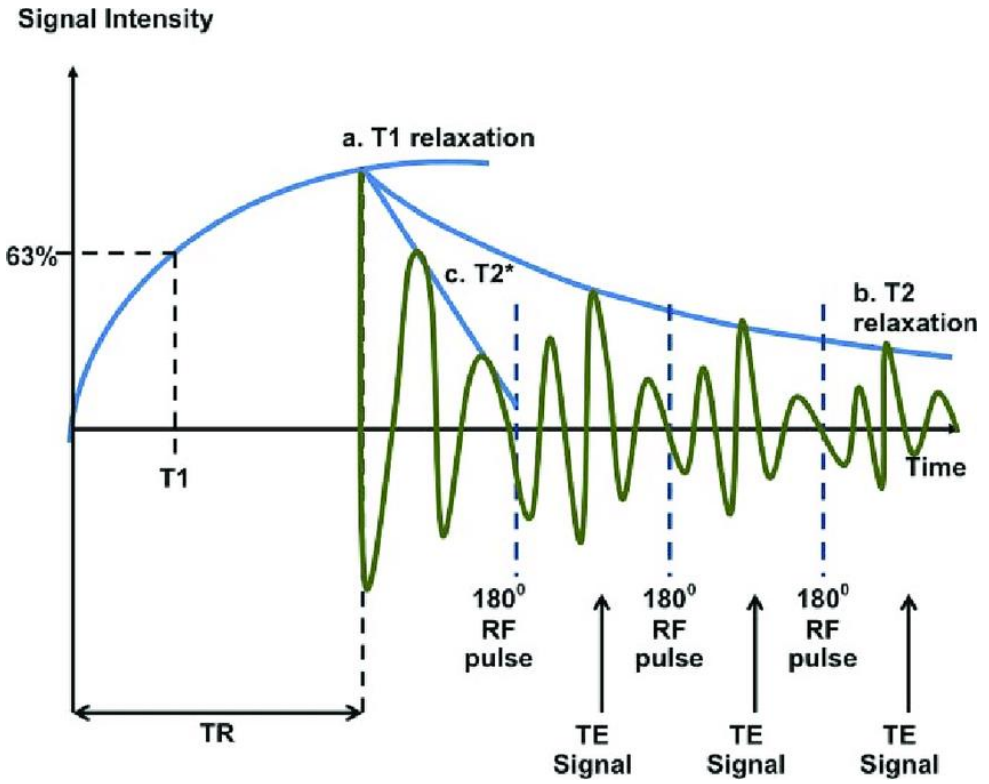


Figure 8: A brief radiofrequency (RF) pulse leads to magnetization that recovers longitudinally with T1. A 180° refocusing pulse is applied to counteract the effect of external static magnetic field (B0) inhomogeneity. The (transversal) decay of the refocusing signal is termed T2. Figure replotted from Currie et al. (Currie et al. 2013)

When imaging at UHF, increases in sensitivity compared to conventional field strengths can be traded for an increase in spatial resolution with still acceptable

signal-to-noise ratios (Polimeni et al. 2018). With this, voxel sizes of ≤ .8mm isotropic are becoming common place (see e.g. Kashyap et al. 2018; Kemper et al. 2015; Kok et al. 2016; Olman et al. 2012; Koopmans et al. 2011). Nevertheless, imaging and data analysis at UHF holds many challenges for which standardized analysis pipelines are still lacking. One fundamental problem is that functional data acquired with echo-planar imaging (EPI) read-out suffers from poor tissue contrast (Gonzalez-Castillo et al. 2013). So, in order to accurately delineate the gray/white matter and gray matter/CSF boundaries necessary for cortical depth sampling, a high tissue contrast anatomical image is required. Such images are typically acquired using either a magnetization prepared rapid gradient echo (MPRAGE, Uğurbil et al. 2003) or a magnetization prepared to rapid acquisition gradient echo (MP2RAGE, Marques et al. 2010) sequence. However, strong distortions in functional images (Jezzard and Clare 1999) acquired at UHF with EPI read-out (due to local magnetic field inhomogeneities) render the alignment to anatomical images, which are less effected by such distortions, difficult. Because distortions occur primarily in the phase encoding direction, we chose to acquire five volumes with opposite phase encoding direction after every functional run to estimate and correct distortions prior to alignment to the anatomical data acquired with MP2RAGE. While distortion correction allows us to render functional data geometrically more similar to anatomical data and therewith aides accurate co-registration, some blurring and loss of spatial resolution due to sub voxel shifts cannot be avoided (Hutton et al. 2002).

Further, despite the technological advancements making high-resolution imaging possible in theory, biological constraints must be considered. Firstly, even with voxel sizes <0.6mm isotropic (Fracasso et al. 2018) histological layers cannot be resolved separately (Amunts, Schleicher, and Zilles 2007). In our sub-millimeter fMRI study, we therefore refer to our data as being cortical depth dependent rather than being specific to laminar layers. Our main region of interest is primary visual cortex, which is relatively flat. Therefore, cortical depth sampling is straightforward compared to highly folded patches of cortex. In highly folded cortex equidistance mesh sampling might lead to biased samples and equivolume sampling strategies are preferred (Kemper et al. 2018). Nevertheless, biological motion causes potential misclassifications of voxels to one or the other cortical depth over time. Even though we scanned well-trained participants that can lay very still for the duration of the scan (>1 hour) to limit the impact of motion artifacts, a residual motion range of ~1mm remains. To correct for the remaining motion effects, we use non-rigid body motion correction algorithms to re-align all images of one run to the first image of the same run. However, subtle changes in magnetic field inhomogeneities introduced by changes in tissue locations within the B_0 field cannot yet be accounted for (Huang et al. 2020).

But the major constraint in blood oxygenation level dependent (BOLD) UHF imaging lies at the very basis of how the BOLD response relates to physiology itself. BOLD fMRI is based on the different magnetic properties of oxygenated and deoxygenated hemoglobin (oxy-HB and deoxy-HB respectively). While oxygenated hemoglobin is diamagnetic, deoxygenated hemoglobin is paramagnetic. Those differences in magnetic susceptibility cause different rates of transversal relaxation (T2 in spin echo and T2* in gradient echo sequences, see figure 9) in surrounding water nuclei (Goebel, Esposito, and Formisano 2006). In effect, the rate of T2 (and T2*) decay is an indicator for the proportion of oxy- to deoxyhemoglobin in the blood at the location measured. The basic assumption of BOLD imaging is that neurons require more oxygen when they are active, and that this oxygen is provided to them by small capillaries, a process termed hemodynamic response. The resulting changes in blood oxygenation can then be measured due to their effect on T2 decay rates (Ogawa et al. 1990). Because the deoxy-HB needs to be transported away from the neurons via larger pial draining veins, the BOLD signal is biased towards cortical depths closer to the pial surface (see e.g. De Martino et al. 2013), especially when using gradient echo, echo planar imaging (GE EPI) as we did in our study. Similar effects are found for the ascending cortical veins that originate at deep levels and increase in diameter towards the pial surface (Duvernoy, Delon, and Vannson 1981). Ascending veins often run radially to the cortical surface and can therefore produce artifacts that are difficult to differentiate from columnar activation patterns (De Martino et al. 2018). Different acquisition sequences such as 3D Grase (Kemper et al. 2015) or VASO (Huber et al. 2015, 2017, 2018) are less susceptible to microvasculature, but at the cost of very small coverage and low sensitivity (De Martino et al. 2013). In our study, we avoid large draining veins by excluding all voxels that could not reliably be mapped in retinotopic space (see methods of chapter 5, pRF mapping). We further use contrasting conditions to restrict interpretations to condition specific differences. However, we currently cannot exclude the possibility that unspecific vasculature driven effects confounded our analysis.

Concluding remarks

The legitimacy of cognitive neuroscience experiments strongly depends on the validity and correct application of methods, be it in the preparation of stimuli, the recording and analysis of data or the statistical interpretation of measurements. The conclusions drawn from the empirical chapters of this thesis make no exception to this rule. While we tried to limit confounding factors where we could, chances are that we overlooked some, and sometimes made the wrong call when faced with the need to make tradeoffs. We will make all analysis code we used freely available with publication of the corresponding chapters to simplify replication of our findings and to facilitate others in finding any errors we might have overlooked.
That being said, our results, in combination with evidence from the literature, suggest that the visual system utilizes redundancies across spatial frequencies to alleviate computational load. Linking the statistical regularities in the environment to brain function has proven to be a fruitful path to pursuing cognitive neuroscience (e.g. Barlow 2001; Friston and Stephan 2007; Ten Oever and Sack 2015). Ultimately, the goal of all brain function is to enable us to make sense of, and meaningfully interact with a world full of statistical patterns.

References

Ales, Justin M, Faraz Farzin, Bruno Rossion, and Anthony M Norcia. 2012. "An Objective Method for Measuring Face Detection Thresholds Using the Sweep Steady-State Visual Evoked Response." *Journal of Vision* 12 (10): 18.

Amunts, Katrin, Axel Schleicher, and Karl Zilles. 2007. "Cytoarchitecture of the Cerebral Cortex—More than Localization." *Neuroimage* 37 (4): 1061–65.

Arsenault, Elizabeth, Ahmad Yoonessi, and Curtis Baker. 2011. "Higher Order Texture Statistics Impair Contrast Boundary Segmentation." *Journal of Vision* 11 (10): 14.

Barlow, Horace. 2001. "The Exploitation of Regularities in the Environment by the Brain." *Behavioral and Brain Sciences* 24 (4): 602–7.

Barone, Pascal, Alexandre Batardiere, Kenneth Knoblauch, and Henry Kennedy. 2000. "Barone, P., Batardiere, A., Knoblauch, K., & Kennedy, H. (2000). Laminar Distribution of Neurons in Extrastriate Areas Projecting to Visual Areas V1 and V4 Correlates with the Hierarchical Rank and Indicates the Operation of a Distance Rule. Journal of Ne." *Journal of Neuroscience* 20 (9): 3263–81.

Bode, Stefan, Daniel Feuerriegel, Daniel Bennett, and Phillip M Alday. 2019. "The Decision Decoding ToolBOX (DDTBOX)–A Multivariate Pattern Analysis Toolbox for Event-Related Potentials." *Neuroinformatics* 17 (1): 27–42.

Bühren, Jens, Evdoxia Terzi, Michael Bach, Wolfgang Wesemann, and Thomas Kohnen. 2006. "Measuring Contrast Sensitivity under Different Lighting Conditions: Comparison of Three Tests." *Optometry and Vision Science* 83 (5): 290–98.

Campbell, Fergus W, and John G Robson. 1968. "Application of Fourier Analysis to the Visibility of Gratings." *The Journal of Physiology* 197 (3): 551.

Coggan, David D, Luke A Allen, Oliver R H Farrar, Andre D Gouws, Antony B Morland, Daniel H Baker, and Timothy J Andrews. 2017. "Differences in Selectivity to Natural Images in Early Visual Areas (V1–V3)." *Scientific Reports* 7 (1): 1–8.

Combrisson, Etienne, and Karim Jerbi. 2015. "Exceeding Chance Level by Chance: The Caveat of Theoretical Chance Levels in Brain Signal Classification and Statistical Assessment of Decoding Accuracy." *Journal of Neuroscience Methods* 250: 126–36.

Cortes, Corinna, and Vladimir Vapnik. 1995. "Support-Vector Networks."

Machine Learning 20 (3): 273–97.
Cox, Michael J, Joanne H Norman, and Peter Norman. 1999. “The Effect of Surround Luminance on Measurements of Contrast Sensitivity.” *Ophthalmic and Physiological Optics* 19 (5): 401–14.
Currie, Stuart, Nigel Hoggard, Ian J Craven, Marios Hadjivassiliou, and Iain D Wilkinson. 2013. “Understanding MRI: Basic MR Physics for Physicians.” *Postgraduate Medical Journal* 89 (1050): 209–23.
Dakin, S C, R F Hess, T Ledgeway, and R L Achtman. 2002. “What Causes Non-Monotonic Tuning of FMRI Response to Noisy Images?” *Current Biology* 12 (14): R476–77.
Dalal, Sarang S, Stefan Rampp, Florian Willomitzer, and Svenja Ettl. 2014. “Consequences of EEG Electrode Position Error on Ultimate Beamformer Source Reconstruction Performance.” *Frontiers in Neuroscience* 8: 42.
de-Wit, Lee, David Alexander, Vebjørn Ekroll, and Johan Wagemans. 2016. “Is Neuroimaging Measuring Information in the Brain?” *Psychonomic Bulletin & Review* 23 (5): 1415–28.
Decharms, R Christopher, and Anthony Zador. 2000. “Neural Representation and the Cortical Code.” *Annual Review of Neuroscience* 23 (1): 613–47.
Douglas, Pamela K, and Ariana Anderson. 2019. “Feature Fallacy: Complications with Interpreting Linear Decoding Weights in FMRI.” In *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*, 363–78. Springer.
Duvernoy, Henri M, SLVJ Delon, and J L Vannson. 1981. “Cortical Blood Vessels of the Human Brain.” *Brain Research Bulletin* 7 (5): 519–79.
Fracasso, Alessio, Peter R Luijten, Serge O Dumoulin, and Natalia Petridou. 2018. “Laminar Imaging of Positive and Negative BOLD in Human Visual Cortex at 7 T.” *NeuroImage* 164: 100–111.
Freeman, Jeremy, Corey M Ziemba, David J Heeger, Eero P Simoncelli, and J Anthony Movshon. 2013. “A Functional and Perceptual Signature of the Second Visual Area in Primates.” *Nature Neuroscience* 16 (7): 974.
Freud, Erez, Jody C Culham, David C Plaut, and Marlene Behrmann. 2017. “The Large-Scale Organization of Shape Processing in the Ventral and Dorsal Pathways.” *Elife* 6: e27576.
Friston, Karl J, and Klaas E Stephan. 2007. “Free-Energy and the Brain.” *Synthese* 159 (3): 417–58.
Goebel, Rainer, Fabrizio Esposito, and Elia Formisano. 2006. “Analysis of

Functional Image Analysis Contest (FIAC) Data with Brainvoyager QX: From Single-subject to Cortically Aligned Group General Linear Model Analysis and Self-organizing Group Independent Component Analysis." *Human Brain Mapping* 27 (5): 392–401.

Goffaux, Valerie, Felix Duecker, Lars Hausfeld, Christine Schiltz, and Rainer Goebel. 2016. "Horizontal Tuning for Faces Originates in High-Level Fusiform Face Area." *Neuropsychologia* 81: 1–11.

Goffaux, Valerie, Judith Peters, Julie Haubrechts, Christine Schiltz, Bernadette Jansma, and Rainer Goebel. 2010. "From Coarse to Fine? Spatial and Temporal Dynamics of Cortical Face Processing." *Cerebral Cortex* 21 (2): 467–76.

Goffaux, Valérie, and Bruno Rossion. 2006. "Faces Are" Spatial"--Holistic Face Perception Is Supported by Low Spatial Frequencies." *Journal of Experimental Psychology: Human Perception and Performance* 32 (4): 1023.

Gonzalez-Castillo, Javier, Kristen N Duthie, Ziad S Saad, Carlton Chu, Peter A Bandettini, and Wen-Ming Luh. 2013. "Effects of Image Contrast on Functional MRI Image Registration." *Neuroimage* 67: 163–74.

Grech, Roberta, Tracey Cassar, Joseph Muscat, Kenneth P Camilleri, Simon G Fabri, Michalis Zervakis, Petros Xanthopoulos, Vangelis Sakkalis, and Bart Vanrumste. 2008. "Review on Solving the Inverse Problem in EEG Source Analysis." *Journal of Neuroengineering and Rehabilitation* 5 (1): 25.

Hansen, Bruce C, and Edward A Essock. 2004. "A Horizontal Bias in Human Visual Processing of Orientation and Its Correspondence to the Structural Components of Natural Scenes." *Journal of Vision* 4 (12): 5.

Haufe, Stefan, Frank Meinecke, Kai Görgen, Sven Dähne, John-Dylan Haynes, Benjamin Blankertz, and Felix Bießmann. 2014. "On the Interpretation of Weight Vectors of Linear Models in Multivariate Neuroimaging." *Neuroimage* 87: 96–110.

Hauk, Olaf. 2004. "Keep It Simple: A Case for Using Classical Minimum Norm Estimation in the Analysis of EEG and MEG Data." *Neuroimage* 21 (4): 1612–21.

Hauk, Olaf, Daniel G Wakeman, and Richard Henson. 2011. "Comparison of Noise-Normalized Minimum Norm Estimates for MEG Analysis Using Multiple Resolution Metrics." *Neuroimage* 54 (3): 1966–74.

Haun, Andrew M, and Edward A Essock. 2010. "Contrast Sensitivity for Oriented Patterns in 1/f Noise: Contrast Response and the Horizontal

Effect." *Journal of Vision* 10 (10): 1.
Haxby, James V. 2012. "Multivariate Pattern Analysis of FMRI: The Early Beginnings." *Neuroimage* 62 (2): 852–55.
Haxby, James V, M Ida Gobbini, Maura L Furey, Alumit Ishai, Jennifer L Schouten, and Pietro Pietrini. 2001. "Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex." *Science* 293 (5539): 2425–30.
Helmholtz, H von. 1853. "Ueber Einige Gesetze Der Vertheilung Elektrischer Ströme in Körperlichen Leitern, Mit Anwendung Auf Die Thierisch-elektrischen Versuche (Schluss.)." *Annalen Der Physik* 165 (7): 353–77.
Hillebrand, Arjan, and Gareth R Barnes. 2003. "The Use of Anatomical Constraints with MEG Beamformers." *Neuroimage* 20 (4): 2302–13.
———. 2005. "Beamformer Analysis of MEG Data." *International Review of Neurobiology* 68: 149–71.
Hincapié, Ana-Sofía, Jan Kujala, Jérémie Mattout, Annalisa Pascarella, Sebastien Daligault, Claude Delpuech, Domingo Mery, Diego Cosmelli, and Karim Jerbi. 2017. "The Impact of MEG Source Reconstruction Method on Source-Space Connectivity Estimation: A Comparison between Minimum-Norm Solution and Beamforming." *Neuroimage* 156: 29–42.
Huang, Pei, Johan D Carlin, Richard N Henson, and Marta M Correia. 2020. "Improved Motion Correction of Submillimetre 7T FMRI Time Series with Boundary-Based Registration (BBR)." *NeuroImage* 210: 116542.
Huber, Laurentius, Jozien Goense, Aneurin J Kennerley, Robert Trampel, Maria Guidi, Enrico Reimer, Dimo Ivanov, Nicole Neef, Claudine J Gauthier, and Robert Turner. 2015. "Cortical Lamina-Dependent Blood Volume Changes in Human Brain at 7 T." *Neuroimage* 107: 23–33.
Huber, Laurentius, Daniel A Handwerker, David C Jangraw, Gang Chen, Andrew Hall, Carsten Stüber, Javier Gonzalez-Castillo, Dimo Ivanov, Sean Marrett, and Maria Guidi. 2017. "High-Resolution CBV-FMRI Allows Mapping of Laminar Activity and Connectivity of Cortical Input and Output in Human M1." *Neuron* 96 (6): 1253–63.
Huber, Laurentius, Dimo Ivanov, Daniel A Handwerker, Sean Marrett, Maria Guidi, Kâmil Uludağ, Peter A Bandettini, and Benedikt A Poser. 2018. "Techniques for Blood Volume FMRI with VASO: From Low-Resolution Mapping towards Sub-Millimeter Layer-Dependent Applications." *Neuroimage* 164: 131–43.

Hutton, Chloe, Andreas Bork, Oliver Josephs, Ralf Deichmann, John Ashburner, and Robert Turner. 2002. “Image Distortion Correction in FMRI: A Quantitative Evaluation.” *Neuroimage* 16 (1): 217–40.

Jezzard, Peter, and Stuart Clare. 1999. “Sources of Distortion in Functional MRI Data.” *Human Brain Mapping* 8 (2-3): 80–85.

Kashyap, Sriranga, Dimo Ivanov, Martin Havlicek, Shubharthi Sengupta, Benedikt A Poser, and Kâmil Uludağ. 2018. “Resolving Laminar Activation in Human V1 Using Ultra-High Spatial Resolution FMRI at 7T.” *Scientific Reports* 8 (1): 1–11.

Kauffmann, Louise, Alan Chauvin, Nathalie Guyader, and Carole Peyrin. 2015. “Rapid Scene Categorization: Role of Spatial Frequency Order, Accumulation Mode and Luminance Contrast.” *Vision Research* 107: 49–57.

Kemper, Valentin G, Federico De Martino, Thomas C Emmerling, Essa Yacoub, and Rainer Goebel. 2018. “High Resolution Data Analysis Strategies for Mesoscale Human Functional MRI at 7 and 9.4 T.” *NeuroImage* 164: 48–58.

Kemper, Valentin G, Federico De Martino, An T Vu, Benedikt A Poser, David A Feinberg, Rainer Goebel, and Essa Yacoub. 2015. “Sub-Millimeter T2 Weighted FMRI at 7 T: Comparison of 3D-GRASE and 2D SE-EPI.” *Frontiers in Neuroscience* 9: 163.

King, J R, and S Dehaene. 2014. “Characterizing the Dynamics of Mental Representations: The Temporal Generalization Method.” *Trends in Cognitive Sciences* 18 (4): 203–10.

Koenderink, Jan J. 1990. “The Brain a Geometry Engine.” *Psychological Research* 52 (2–3): 122–27.

Kok, Peter, Lauren J Bains, Tim van Mourik, David G Norris, and Floris P de Lange. 2016. “Selective Activation of the Deep Layers of the Human Primary Visual Cortex by Top-down Feedback.” *Current Biology* 26 (3): 371–76.

Koopmans, Peter J, Markus Barth, Stephan Orzada, and David G Norris. 2011. “Multi-Echo FMRI of the Cortical Laminae in Humans at 7 T.” *Neuroimage* 56 (3): 1276–85.

Lin, Fa-Hsuan, Thomas Witzel, Seppo P Ahlfors, Steven M Stufflebeam, John W Belliveau, and Matti S Hämäläinen. 2006. “Assessing and Improving the Spatial Accuracy in MEG Source Localization by Depth-Weighted Minimum-Norm Estimates.” *Neuroimage* 31 (1): 160–71.

Malach, Rafael, J B Reppas, R R Benson, K K Kwong, H Jiang, W A

Kennedy, P J Ledden, T J Brady, B R Rosen, and R B Tootell. 1995. "Object-Related Activity Revealed by Functional Magnetic Resonance Imaging in Human Occipital Cortex." *Proceedings of the National Academy of Sciences* 92 (18): 8135–39.

Margalit, Eshed, Manan P Shah, Bosco S Tjan, Irving Biederman, Brenton Keller, and Rorry Brenner. 2016. "The Lateral Occipital Complex Shows No Net Response to Object Familiarity." *Journal of Vision* 16 (11): 3.

Markov, Nikola T, Julien Vezoli, Pascal Chameau, Arnaud Falchier, René Quilodran, Cyril Huissoud, Camille Lamy, Pierre Misery, Pascale Giroud, and Shimon Ullman. 2014. "Anatomy of Hierarchy: Feedforward and Feedback Pathways in Macaque Visual Cortex." *Journal of Comparative Neurology* 522 (1): 225–59.

Marques, José P, Tobias Kober, Gunnar Krueger, Wietske van der Zwaag, Pierre-François Van de Moortele, and Rolf Gruetter. 2010. "MP2RAGE, a Self Bias-Field Corrected Sequence for Improved Segmentation and T1-Mapping at High Field." *Neuroimage* 49 (2): 1271–81.

Martino, Federico De, Essa Yacoub, Valentin Kemper, Michelle Moerel, Kâmil Uludağ, Peter De Weerd, Kamil Ugurbil, Rainer Goebel, and Elia Formisano. 2018. "The Impact of Ultra-High Field MRI on Cognitive and Computational Neuroimaging." *Neuroimage* 168: 366–82.

Martino, Federico De, Jan Zimmermann, Lars Muckli, Kamil Ugurbil, Essa Yacoub, and Rainer Goebel. 2013. "Cortical Depth Dependent Functional Responses in Humans at 7T: Improved Specificity with 3D GRASE." *PloS One* 8 (3).

Meng, Ming, Tharian Cherian, Gaurav Singal, and Pawan Sinha. 2012. "Lateralization of Face Processing in the Human Brain." *Proceedings of the Royal Society B: Biological Sciences* 279 (1735): 2052–61.

Mu, Tonglin, and Sheng Li. 2013. "The Neural Signature of Spatial Frequency-Based Information Integration in Scene Perception." *Experimental Brain Research* 227 (3): 367–77. https://doi.org/10.1007/s00221-013-3517-1.

Murray, Scott O, Daniel Kersten, Bruno A Olshausen, Paul Schrater, and David L Woods. 2002. "Shape Perception Reduces Activity in Human Primary Visual Cortex." *Proceedings of the National Academy of Sciences* 99 (23): 15164–69.

Neymotin, Samuel A, Dylan S Daniels, Blake Caldwell, Robert A

McDougal, Nicholas T Carnevale, Mainak Jas, Christopher I Moore, Michael L Hines, Matti Hämäläinen, and Stephanie R Jones. 2020. "Human Neocortical Neurosolver (HNN), a New Software Tool for Interpreting the Cellular and Network Origin of Human MEG/EEG Data." *Elife* 9.

Oever, Sanne Ten, and Alexander T Sack. 2015. "Oscillatory Phase Shapes Syllable Perception." *Proceedings of the National Academy of Sciences* 112 (52): 15833–37.

Ogata, Katsuhiko. 1995. *Discrete-Time Control Systems*. Vol. 2. Prentice Hall Englewood Cliffs, NJ.

Ogawa, Seiji, Tso-Ming Lee, Alan R Kay, and David W Tank. 1990. "Brain Magnetic Resonance Imaging with Contrast Dependent on Blood Oxygenation." *Proceedings of the National Academy of Sciences* 87 (24): 9868–72.

Olman, Cheryl A, Noam Harel, David A Feinberg, Sheng He, Peng Zhang, Kamil Ugurbil, and Essa Yacoub. 2012. "Layer-Specific FMRI Reflects Different Neuronal Computations at Different Depths in Human V1." *PloS One* 7 (3).

Olshausen, Bruno A, and David J Field. 1997. "Sparse Coding with an Overcomplete Basis Set: A Strategy Employed by V1?" *Vision Research* 37 (23): 3311–25.

———. 2004. "Sparse Coding of Sensory Inputs." *Current Opinion in Neurobiology* 14 (4): 481–87.

Peyrin, Carole, Martial Mermillod, Sylvie Chokron, and Christian Marendaz. 2006. "Effect of Temporal Constraints on Hemispheric Asymmetries during Spatial Frequency Processing." *Brain and Cognition* 62 (3): 214–20.

Polimeni, Jonathan R, Ville Renvall, Natalia Zaretskaya, and Bruce Fischl. 2018. "Analysis Strategies for High-Resolution UHF-FMRI Data." *Neuroimage* 168: 296–320.

Popov, Tzvetan, Robert Oostenveld, and Jan M Schoffelen. 2018. "FieldTrip Made Easy: An Analysis Protocol for Group Analysis of the Auditory Steady State Brain Response in Time, Frequency, and Space." *Frontiers in Neuroscience* 12.

Rainer, Gregor, Mark Augath, Torsten Trinath, and Nikos K Logothetis. 2002. "The Effect of Image Scrambling on Visual Cortical BOLD Activity in the Anesthetized Monkey." *Neuroimage* 16 (3): 607–16.

Riesenhuber, Maximilian, and Tomaso Poggio. 2002. "Neural Mechanisms of Object Recognition." *Current Opinion in Neurobiology* 12 (2): 162–

68.
Rossion, Bruno, Carrie A Joyce, Garrison W Cottrell, and Michael J Tarr. 2003. “Early Lateralization and Orientation Tuning for Face, Word, and Object Processing in the Visual Cortex.” *Neuroimage* 20 (3): 1609–24.
Sadr, Javid, and Pawan Sinha. 2004. “Object Recognition and Random Image Structure Evolution.” *Cognitive Science* 28 (2): 259–87.
Sekihara, Kensuke, Maneesh Sahani, and Srikantan S Nagarajan. 2005. “Localization Bias and Spatial Resolution of Adaptive and Non-Adaptive Spatial Filters for MEG Source Reconstruction.” *Neuroimage* 25 (4): 1056–67.
Shannon, Claude Elwood. 1948. “A Mathematical Theory of Communication.” *Bell System Technical Journal* 27 (3): 379–423.
Solomon, Chris, and Toby Breckon. 2011. *Fundamentals of Digital Image Processing: A Practical Approach with Examples in Matlab*. John Wiley & Sons.
Stojanoski, Bobby, and Rhodri Cusack. 2014. “Time to Wave Good-Bye to Phase Scrambling: Creating Controlled Scrambled Images Using Diffeomorphic Transformations.” *Journal of Vision* 14 (12): 6.
Uğurbil, Kâmil, Gregor Adriany, Peter Andersen, Wei Chen, Michael Garwood, Rolf Gruetter, P-G Henry, S-G Kim, Haiying Lieu, and Ivan Tkac. 2003. “Ultrahigh Field Magnetic Resonance Imaging and Spectroscopy.” *Magnetic Resonance Imaging* 21 (ARTICLE): 1263–81.
Varoquaux, Gaël, Pradeep Reddy Raamana, Denis A Engemann, Andrés Hoyos-Idrobo, Yannick Schwartz, and Bertrand Thirion. 2017. “Assessing and Tuning Brain Decoders: Cross-Validation, Caveats, and Guidelines.” *NeuroImage* 145: 166–79.
Veen, Barry D Van, Wim Van Drongelen, Moshe Yuchtman, and Akifumi Suzuki. 1997. “Localization of Brain Electrical Activity via Linearly Constrained Minimum Variance Spatial Filtering.” *IEEE Transactions on Biomedical Engineering* 44 (9): 867–80.
Vinje, William E, and Jack L Gallant. 2000. “Sparse Coding and Decorrelation in Primary Visual Cortex during Natural Vision.” *Science* 287 (5456): 1273–76.
Vlamings, Petra H J M, Valerie Goffaux, and Chantal Kemner. 2009. “Is the Early Modulation of Brain Activity by Fearful Facial Expressions Primarily Mediated by Coarse Low Spatial Frequency Information?” *Journal of Vision* 9 (5): 12.

Williams, Mark A, Sabin Dang, and Nancy G Kanwisher. 2007. “Only Some Spatial Patterns of FMRI Response Are Read out in Task Performance.” *Nature Neuroscience* 10 (6): 685–86.

Wolters, Carsten Hermann, Alfred Anwander, Burkhard Maess, R S Macleod, and Angela D Friederici. 2004. “The Influence of Volume Conduction Effects on the EEG/MEG Reconstruction of the Sources of the Early Left Anterior Negativity.” In *The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2:3569–72. IEEE.

Publications and Presentations

Articles published in per-reviewed journals

Jacobs, C., **Petras, K**., Moors, P., & Goffaux, V. (2020). Contrast versus identity encoding in the face image follow distinct orientation selectivity profiles. *PloS one*, *15*(3), e0229185.

Petras, K., ten Oever, S., Jacobs, C., & Goffaux, V. (2019). Coarse-to-fine information integration in human vision. *NeuroImage*, *186*, 103-112.

Petras, K., ten Oever, S., & Jansma, B. M. (2016). The effect of distance on moral engagement: event related potentials and alpha power are sensitive to perspective in a virtual shooting task. *Frontiers in psychology*, *6*, 2008.

Oral presentations at conferences with scientific selection committee

K. Petras, S. S. Dalal, V. Goffaux (2019). Image coherence modulates strength of feedback to early visual cortex. European conference on visual perception, Leuven, Belgium

K. Petras, S. ten Oever, C. Jacobs, V. Goffaux (2018). Coarse image information guides the integration of fine details (if you let it). European conference on visual perception, Trieste, Italy

K. Petras, L. Geers, V. Goffaux (2016). Vertically oriented cues to face identification are susceptible to color manipulations. Annual meeting of the Belgian Association of Psychological Sciences, Antwerp, Belgium

Poster presentations at conferences with scientific selection committee

K. Petras, V. Goffaux (2017). Tracking spatial frequency integration with EEG. European conference on visual perception, Berlin, Germany

K. Petras, V. Goffaux (2017). EEG classification to investigate the roles of high and low spatial frequencies in full spectrum image processing. International conference of cognitive neuroscience, Amsterdam, The Netherlands

K. Petras, V. Goffaux (2017). Tracking spatial frequency in EEG source-space. Annual meeting of the Belgian society for neuroscience, Ghent, Belgium

K. Petras, V. Goffaux (2017). Tracking spatial frequency in EEG source-space. Cutting EEG conference, Glasgow, UK

K. Petras, L. Geers, V. Goffaux (2016). Vertically oriented cues to face identification are susceptible to color manipulations. European conference on visual perception, Barcelona, Spain