

Human Face Processing: Recurrent Interactions Across the Visual System

Human Face Processing: Recurrent Interactions Across the Visual System

Jolien P. SCHUURMANS

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

Promoter

Prof. Dr. Valérie GOFFAUX (UCLouvain, Belgium)

Accompanying committee

Prof. Dr. Olivier COLLIGNON (UCLouvain, Belgium)

Prof. Dr. Hans OP DE BEECK (KULeuven, Belgium)

Collection de thèses de l'Université catholique de Louvain, 2023

Cover generated by OpenAI's DALL-E, accessed November 2023

President

Prof. Dr. Olivier COLLIGNON (UCLouvain, Belgium)

Accompanying committee and jury

Prof. Dr. Valérie GOFFAUX (UCLouvain, Belgium)

Prof. Dr. Hans OP DE BEECK (KULeuven, Belgium)

Prof. Dr. Louise KAUFFMANN (l'Université Grenoble Alpes, France)

Prof. Dr. Katharina DOBS (Justus-Liebig Universität Giessen, Germany)

Prof. Dr. Virginie CROLLEN (UCLouvain, Belgium)

Abstract

This thesis examines the role of recurrent processing in the human visual system, with a special emphasis on the perception of complex stimuli, notably faces. It draws inspiration from theories proposing that recurrent processing operates in a coarse-to-fine manner, where initial processing of coarse/low spatial frequencies (LSFs) guides subsequent processing of finer/high spatial frequencies (HSFs). Additionally, the need for recurrent processes increases when stimuli are degraded. The thesis is structured around three primary research questions.

The first study (Chapter 2) investigated the role of the primary visual cortex (V1) in coarse-to-fine processing. Leaning on the conception of V1 as an 'active blackboard' that integrates visual information in a coarse-to-fine manner. The study employed fMRI and applied backward masking over time and of different spatial frequency (SF) ranges in facial stimuli to interrupt LSF or HSF processing. Results showed strong disruptions in LSF processing as a result of early masking, reducing over time, whereas HSF processing was affected more at later stages. These effects were not only observed in V1 but also in ventral, dorsal, and orbitofrontal brain regions. The findings were in line with existing coarse-to-fine processing theories.

The second thread of inquiry (Chapter 3) examined the behavioural outcomes of disrupting coarse-to-fine processing. While the results affirmed that later masking proves less disruptive for identity recognition, no effects from the SF manipulation of the mask were detected. This finding suggests the possibility of a transient coarse-to-fine processing window, potentially too brief for observable behavioural responses.

The third study (Chapter 4), focussing on the relationship between stimulus degradation and recurrent processing, showed that noisy stimuli necessitate more recurrent processing. In V1/V2, the impact of masking was reduced with time, regardless of the stimulus degradation level, indicating that extended recurrent processes allows for more suppressive feedback from higher-level areas. Moreover, backward masking effects were more prominent with degraded stimuli, which naturally demand increased recurrent processing. Higher-order face-selective regions showed increased responses with prolonged stimulus exposure and clear face stimuli. The latter is in line with the regions' tendency to respond more robustly to apparent faces.

In summary, this research provides insights into the mechanics and role of recurrent processing in perception of complex stimuli. It underscores the function of coarse-to-fine processing in V1 during shaping visual perception and the impact of varying stimulus degradation levels on recurrent processes. By consolidating findings from diverse brain regions and behavioural outcomes, this study enhances the understanding of real-world visual perception mechanisms.

Résumé

Cette thèse examine le rôle du traitement récurrent dans le système visuel humain, avec un accent particulier sur la perception de stimuli complexes, notamment les visages. Elle s'inspire de théories proposant que le traitement récurrent fonctionne de manière grossière à fine, où le traitement initial des fréquences spatiales basses/grossières (FSB) guide le traitement ultérieur des fréquences spatiales plus fines/élevées (FSE). De plus, le besoin de processus récurrents augmente lorsque les stimuli sont dégradés. La thèse est structurée autour de trois questions de recherche principales.

La première étude (Chapitre 2) a investigué le rôle du cortex visuel primaire (V1) dans le traitement grossier à fin. S'appuyant sur la conception de V1 comme un 'tableau noir actif' qui intègre les informations visuelles de manière grossière à fine, l'étude a employé l'IRMf et appliqué le masquage postérieur sur plusieurs gammes de fréquences spatiales (FS) dans des stimuli de visages pour interrompre le traitement FSB ou FSE. Les résultats ont montré des perturbations importantes dans le traitement FSB avec le masquage précoce, qui diminuent avec le temps, tandis que le traitement FSE était plus affecté aux étapes ultérieures. Ces effets ont été observés non seulement dans V1 mais aussi dans les régions cérébrales ventrales, dorsales et orbitofrontales. Les résultats étaient conformes aux théories existantes du traitement grossier à fin.

Le deuxième fil d'enquête (Chapitre 3) a examiné les résultats comportementaux de la perturbation du traitement grossier à fin. Nos résultats ont montré qu'un masque tardif s'avère moins perturbateur pour la reconnaissance de l'identité et aucun effet de la manipulation de la FS du masque n'a non plus été détecté. Cette constatation suggère la possibilité d'une fenêtre de traitement grossier à fin transitoire, potentiellement trop brève pour des réponses comportementales observables.

La troisième étude (Chapitre 4), se concentrant sur la relation entre la dégradation des stimuli et le traitement récurrent, a montré que les stimuli bruités nécessitent plus de traitement récurrent. Dans V1/V2, l'impact du masquage a été réduit avec le temps, indépendamment du niveau de dégradation du stimulus, indiquant que les processus récurrents prolongés permettent un retour suppressif plus important des zones de niveau supérieur. De plus, les effets du masquage postérieur étaient plus prononcés avec des stimuli dégradés, qui demandent naturellement un traitement récurrent élevé. Les régions sélectives du visage de haut ordre ont montré des réponses accrues avec une exposition prolongée aux stimuli ainsi qu'avec des stimuli faciaux clairs. Ce dernier point est en accord avec la tendance des régions à répondre plus robustement aux visages apparents.

En résumé, cette recherche fournit un aperçu de la mécanique du rôle du traitement récurrent dans la perception de stimuli complexes. Elle souligne la fonction du traitement grossier à fin dans V1 dans la perception visuelle ainsi que l'impact des niveaux de dégradation des stimuli variables sur les processus

récurrents. En consolidant les résultats de diverses régions cérébrales et des résultats comportementaux, cette étude améliore la compréhension des mécanismes de perception visuelle du monde réel.

Acknowledgements

I would like to express gratitude to the following individuals who have played an important role in my journey through this PhD, each bringing their unique presence and support.

First and foremost, I would like to express my gratitude to Valérie Goffaux for not only agreeing to supervise me and giving me feedback on my work, but also for listening to mine in return. I always felt welcome in your office, whether I needed assistance with work-related questions, had something important to discuss, or simply wanted to have a casual chat. And let's not forget the introduction to your cat, dog, and sheep!

I would also like to extend my thanks to the accompanying committee members, Olivier Collignon and Hans op de Beeck, for their continuous support and for monitoring my progress throughout my PhD. A special appreciation goes out to Louise Kauffmann, Katharina Dobs, and Virginie Crollen for joining the doctoral jury and dedicating their time and effort to reading and evaluating this thesis. Your thorough examination and valuable insights are appreciated.

Serge Dumoulin, thank you for welcoming me into your lab. I want to extend my appreciation to all the members of the Spinoza Centre For Neuroimaging for assisting me in acquiring 7T MRI data, guiding me through the analysis process, and, of course, teaching me how to play Fussball properly. Because of you, we even bought a tiny Fussball table at the Goffauxlab.

A special shout-out to Umut for accompanying me throughout this journey. We did it! Umut, you will forever hold a special place in my heart as my favourite ~~Greek~~ Turkish colleague. And don't forget: an apple a day might keep the doctorate away, but two bananas will make sure you never cramp. Hélène, thank you for joining the office space during our time of writing. Your presence provided that much-needed social control that helped both Umut and me concentrate better. I also have endless gratitude for you helping me with the French abstract on the day of my submission, and of course the socks you gave me! Of course, thanks to the rest of my (ex) lab colleagues, Mrittika, Alexia, Yu-Fang, Christiane, Marius and Ece. For making sure there was always someone to talk to in the lab, either for a casual chat or to listen to me complaining about something I was stuck with (and then help). And especially Mrittika, for not complaining about the time I accidentally shut off the power around the MRI scanner! Matthew, you deserve a special mention for introducing me to the delightful world of crisp sandwiches and bread and butter pudding. Your willingness to take care of our cat while we were away and enduring those endless hours at the hospital scanner with me during the Christmas weekends (all while munching on those crisp sandwiches) means the world to me. Kirsten, your guidance during the early years of my PhD was invaluable. And it wasn't just about knowledge; you were always there for a good conversation accompanied by Marmorkuchen.

Hidde, I must admit that it was challenging for me at the beginning of my PhD to work without your Labrador-vibe around. And I'm glad we stayed in touch. Thank you for allowing me to scan your brain again and for the random Skype sessions we had. A special thanks goes out to: Jolijn, Marise, Gail, and Kelsey. You have patiently listened to countless hours of voice memos, and I am grateful for your unwavering support. Jolijn, your Sinterklaas packages brought me so much joy, making me feel connected to Dutch holidays. Plus, your MRI data looked fantastic!

Mam, ik wil je graag bedanken. Van samen wandelen tot telefoongesprekken die altijd net even langer duren dan gepland, je bent er altijd voor me. En pap, dankje dat je altijd geïnteresseerd bent in mijn loopbaan als we deze bespreken tijdens een van onze bezoeken aan restaurants in de buurt. Daarnaast was het erg fijn om een vakantieadresje te hebben gehad in Virginia Beach. Maar ik ben nu vooral blij dat je Ingeborg en Ellis weer dichterbij hebt gebracht. Inge en Paul, het is niet altijd gemakkelijk voor me geweest om zo verweg van jullie te wonen, zeker tijdens de gelukkige en zware momenten in de afgelopen jaren. Maar wanneer we samen zijn is het meteen als van ouds en daar ben ik erg dankbaar voor. Ook wil ik erg graag mijn schoonouders Margo en Loes bedanken. Erg fijn dat ik bij jullie kon verblijven toen ik in Nederland moest zijn voor werk (of plezier). Ik heb de afgelopen winter veel steun aan jullie gehad! En natuurlijk Corien en Astrid, jullie staan altijd voor mij klaar wanneer ik taal gerelateerde vragen heb of om advies te geven. Dankjulliewel. Julia en Lisa, bedankt voor het geduldig verdragen van mijn verzoeken om zo stil mogelijk te liggen in de 7T-scanner. Jullie data zijn natuurlijk van onschatbare waarde gebleken voor mijn onderzoek. And, a special mention to Nicole, whom I know I can call in the middle of the night if I need to. Thank you for always being there for me.

And last but certainly not least, my partner Djimo. Thank you for joining me on this adventure, moving with me to Belgium, and even marrying me along the way. You have been *my rock*, my confidant, and a huge source of motivation for me. Whether acting as my rubber duck when I was stuck, providing me with Cécémel during intense work days, or accompanying me on our daily walks around the lake, you have been an indispensable part of my success. I could not have completed this thesis without you, as you have been a valuable participant in all my studies.

I would like to honour the memory of Nikita, my little sister, who couldn't be here to witness the end of this journey.

Contents

Abstract	v
Résumé	vii
Acknowledgements	ix
1 General introduction	1
Visual perception	1
Image properties	1
Filtering spatial frequencies	2
Information processing in the brain	4
Hierarchical organisation	4
Recurrent processing	8
Degraded stimuli	13
Coarse-to-fine processing	14
Backward masking	17
Backward masking to study recurrent processing	17
Masking coarse-to-fine	20
Limitations of previous studies	20
Faces as complex stimuli in research	21
Face processing in early visual areas	23
Outline of the thesis	24
Methodology	25
Functional Magnetic Resonance Imaging	25
General Linear Model	27
Psychophysiological interaction analysis	28
Linear Mixed effects Modelling analysis	28
2 Backward masking reveals coarse-to-fine dynamics in human V1	31
Abstract	31
Introduction	32
Methods	36
Subjects	36
Stimuli	36
Main experiment	36
Functional localiser	39
Procedure	39
Main experiment	39
Functional localiser	41
MRI acquisition	41

Preprocessing	41
Anatomical data pre-processing	41
Functional data pre-processing	42
Regions of interest	43
Statistical Analyses	45
Behaviour analysis	45
ROI analysis	45
Whole-volume group-level analysis	46
Psychophysiological Interaction	47
Results	48
Behavioural results	48
ROI results	49
V1	49
FFA	51
Whole-volume group-level results	52
Psychophysiological Interaction	55
Discussion	56
Supplementary Material	63
Supplementary Material 1: Functional localiser stimuli	63
Supplementary Material 2: Different V1 delineation	64
Supplementary Material 3: Plots of beta values from GLM	67
Supplementary Material 4: Linear Mixed effect Models: stepwise variable selection	68
Supplementary Material 5: Tables with fixed effects	69
Supplementary Material 6: Extra figures: V2 and OFA	71
3 The role of coarse-to-fine integration in human face identity recognition	73
Abstract	73
Introduction	74
Methods	78
Subjects	78
Stimuli	78
Procedure	80
Analyses	81
Psychometric fitting	81
Linear Mixed effects Modeling	82
Results	84
Psychometric fitting	84
Linear Mixed effects Modelling results	84
Discussion	86

4	Enhanced recurrent processing under conditions of visual input degradation	93
	Abstract	93
	Introduction	94
	Methods	97
	Subjects	97
	Stimuli	98
	Main experiment	98
	Functional localiser	100
	Procedure	100
	SNR test	101
	Main experiment	101
	Functional localiser	103
	MRI acquisition	103
	Preprocessing	104
	Anatomical data	104
	Functional data	105
	Regions of interest	106
	Statistical analyses	108
	Behavioural analysis	108
	Region of interest analysis	108
	Whole brain group-level analysis	110
	Correlation analysis	110
	Psychophysiological Interaction	111
	Results	111
	Behavioural results	111
	Region of interest results	113
	V1/V2	113
	FFA	114
	OFA	114
	Correlation results	118
	V1/V2 - FFA	118
	V1/V2 - OFA	118
	OFA - FFA	119
	Discussion	121
	Supplementary materials	127
5	General discussion	129
	Summary	129
	Studying coarse-to-fine integration with full-spectrum stimuli	131
	V1 as an "active blackboard"	134
	Investigating recurrent processing	134
	Coarse-to-fine processing in behaviour	137
	Simplified investigations into coarse-to-fine integration	138
	Predictive coding framework	140

Statistical Linear Mixed effects Modelling approach	142
Timing in the visual system	144
Face stimuli in research	146
Beyond faces	148
Concluding remarks	149
Publications and Presentations	151
Articles published in peer-reviewed journals	151
Conferences with a scientific selection committee	151
Oral	151
Poster	151
References	153

List of Figures

1.1	Spectral analysis of spatial frequencies in greyscale images	3
1.2	Depiction of dorsal and ventral cortical pathways	7
1.3	Laminar structure of V1 in a macaque monkey	9
1.4	Kanizsa triangle and fMRI responses	12
1.5	Proposed model of integrated visual processing by Bullier	15
1.6	Disruption of recurrent interactions by masking stimuli	19
1.7	Model of the hemodynamic response function.	26
2.1	Expected responses of V1	35
2.2	Stimulus generation	38
2.3	Design of main experiment	40
2.4	V1 and FFA ROI selection	44
2.5	Results in V1 and FFA	50
2.6	Whole-volume group-level analysis	53
2.7	PPI analysis: V1-FFA	55
2.8	Supplementary: Functional localiser stimuli	63
2.9	Supplementary: Results in differently defined V1	65
2.10	Supplementary: Results in both hemispheres of V1 and FFA	67
2.11	Supplementary: Results in V2 and OFA	71
2.12	Supplementary: PPI analysis - V1-FFA full model plot	72
3.1	Depiction of spatial frequencies in a photograph of a crab	74
3.2	Face blending example and trial structure	79
3.3	Example of staircase and psychometric function	82
3.4	Logistic psychometric fitting all subjects	83
3.5	Results for stimulus signal-to-noise thresholds	85
4.1	Alpha blending example	99
4.2	Experimental design	102
4.3	Functionally defined V1/V2, FFA and OFA	107
4.4	Behavioural results	112
4.5	Results V1/V2	115
4.6	Results FFA	116
4.7	Results OFA	117
4.8	Correlation analysis results	120
4.9	Results in V1/V2 and FFA incl scrambled condition	127
5.1	Interrelated spatial frequency scales	132
5.2	Model of contextual facilitation in object recognition by Bar	135
5.3	Monochrome image	141
5.4	Colourised version of pervious figure	143

5.5	Response latencies across various cortical areas	145
-----	--	-----

List of Tables

2.1	ROI size and coordinates	45
2.2	ANOVA behaviour, V1 and FFA	48
2.3	Model predictions	51
2.4	Whole-volume group-level analysis	54
2.5	Supplementary: ANOVA, differently defined V1	66
2.6	Supplementary: Model predictions differently defined V1	66
2.7	Supplementary: Fixed effects differently defined V1	66
2.8	Supplementary: LMM stepwise variable selection	68
2.9	Supplementary: Accuracy Fixed effects	69
2.10	Supplementary: V1 Fixed effects	69
2.11	Supplementary: FFA Fixed effects	69
2.12	Supplementary: PPI Fixed effects	70
2.13	Supplementary: V2 Fixed effects	70
2.14	Supplementary: OFA Fixed effects	70
3.1	Model predictions	85
4.1	ROI size and coordinates	106
4.2	Correlation analysis results	119

List of Abbreviations

3T	3Tesla
7T	7Tesla
ANOVA	Analysis Of Variance
BA	Brodmann's area
BIC	Bayesian Information Criterion
BIDS	Brain Imaging Data Structure
BOLD	Blood-Oxygen Level-Dependent
CI	Confidence Interval
cpd	cycles per degree
cpf	cycles per face width
CSF	Cerebrospinal Fluid
DLPFC	Dorsolateral Prefrontal Cortex
DenDF	Denominator Degrees of Freedom
DTI	Diffusion Tensor Imaging
EBA	Extrastriate Body Area
EEG	Electroencephalography
EPI	Echo Planar Imaging
ERP	Event-Related Potential
FA	Flip Angle
FD	Framewise Displacement
FDR	False Discovery Rate
FFA	Fusiform Face Area
fMRI	functional Magnetic Resonance Imaging
FOV	Field Of View
FrACT	Freiburg Visual Acuity and Contrast Test
FWHM	Full Width at Half Maximum
GLM	General Linear Model
GM	Grey-Matter
hemi	hemisphere
HRF	Hemodynamic Response Function
HSF	High Spatial Frequency
HSV	Hue Saturation Value
Hz	Hertz
IPS	Intraparietal Sulcus
IT	Inferotemporal
LGN	Lateral Geniculate Nucleus
LMM	Linear Mixed effects Model
LOC	Lateral Occipital Complex
LSF	Low Spatial Frequency
M-pathway	Magnocellular-pathway

mafc	multi-alternative forced-choice
MEG	Magnetoencephalography
M	Mean
MRI	Magnetic Resonance Imaging
ms	milliseconds
MST	Medial Superior Temporal area
MT	Mid Temporal area or V5
NORDIC	Noise Reduction Distribution Correction
NumDF	Numerator Degrees of Freedom
OFA	Occipital Face Area
P-pathway	Parvocellular-pathway
pFC	prefrontal Cortex
PPC	Posterior Parietal Cortex
PPA	Parahipocampal Place Area
PPI	Psychophysiological Interaction
RMS	Root Mean Square
ROI	Region Of Interest
RT	Response Time
RSVP	Rapid Serial Visual Presentation
SD	Standard Deviation
SE	Standard Error
SF	Spatial Frequency
SNR	Signal-to-Noise Ratio
STS	Superior Temporal Sulcus
SOA	Stimulus Onset Asynchrony
T1w	T1-weighted
T2w	T2-weighted
TE	Echo Time
TI	Inversion Time
TMS	Transcranial Magnetic Stimulation
TR	Repetition Time
tSNR	temporal Signal-to-Noise Ratio
V1	Primary Visual area
V2	Secondary Visual area
VWFA	Visual Word Form Area
WM	White-Matter

Chapter 1

General introduction

Visual perception

Our everyday interactions largely depend on our ability to interpret the wide range of sensory input we encounter, with visual perception being one of the most crucial mechanisms. From simple light signals to complex scenes, the brain's visual processing system continuously interprets and makes sense of our surroundings. This process allows us to translate complex visual stimuli into meaningful experiences by turning lights and shadows into recognisable shapes and objects, thus helping us make sense of our reality. How does the brain transform simple light signals into intricate visual experiences? This question has intrigued scientists for decades. The intricacies of visual perception, involving a complex interplay of brain regions, remains a subject of ongoing research.

Image properties

To be able to understand visual perception, it is important to consider the properties of the input processed by the visual system. Among these properties, spatial frequencies (SFs) are essential components of images and significantly affect how we perceive and recognise things around us. SFs pertain to the spatial scales of luminance intensity variations within an image. The top row of Figure 1.1 depicts simple black and white striped images. If the stripes are wide and spread out (left image), we say the image has low SF (LSF); the changes in intensity are gradual across the image. If the stripes are slender and tightly clustered (right image), the image has high SF (HSF); indicative of sharp transitions in light intensity. Thus, SF is a measure of how rapidly the light intensity changes in image space and corresponds to the number of cycles of luminance variation per unit of distance.

Complex images, such as faces or scenes (see middle row of Figure 1.1), are composed of a combination of different SFs. LSFs outline the general shape of a face or the layout of a scene, while HSFs represent the details such as individual facial features or leaves on a tree. The spectrum of SFs composing an image generally exhibits strong spatial correlation; with both low and high SFs contributing to provide a comprehensive visual representation of our surroundings. The distribution of SF in these images is not evenly distributed. A scene or face stimulus typically contains more LSF components than HSF components,

following an inverse power law, called the $1/f^\alpha$ rule (with α around 1.2 for natural scenes, and closer to 2 for face images). For faces, this implies that as the SF doubles, the power (or energy) in that range is halved. This reflects the fact that in naturalistic visual stimuli, coarse-scale/LSF features are more prominent than fine-scale/HSF details. Furthermore, images depicting natural scenes often exhibit a dominance of cardinal directions, namely the horizontal and vertical orientations (Coppola et al., 1998; Torralba & Oliva, 2003). This pattern of statistical similarities is even more pronounced within specific categories of visual objects. For example, in the case of facial images, there are only minor differences in spectral properties among different examples, with the maximum contrast being evident in the vertical configuration of horizontally extended features such as the eyes or the mouth (Dakin & Watt, 2009).

Filtering spatial frequencies

An approach to analyse the SFs of an image is the application of the Fourier transform (Brigham, 1988; Cochran et al., 1967). This technique mathematically reinterprets an image into its constituent sine and cosine waves, each representing a specific SF range. In the spectral representation of an image, each point in the frequency domain corresponds to a specific SF in the original image, with the magnitude (amplitude, or contrast) at each point indicating the contribution of that frequency to the overall image (Figure 1.1, bottom row). By translating the image into the frequency domain, an image can be summarised by its amplitude spectrum and phase spectrum. The amplitude spectrum provides the distribution of amplitude as a function of SFs and orientations. The phase spectrum characterises the relative position of SFs in space.

In the frequency (or Fourier) domain, we are able to manipulate specific SF ranges through filtering. This involves the multiplication of the image's frequency spectrum by a specific filter matrix, which is designed to allow certain frequencies to pass through while attenuating others. For instance, the Gaussian filter, which has a bell-shaped curve in the frequency domain, smoothly transitions from low to high frequencies up to a certain cut-off point, beyond which it rapidly approaches zero, thus gently blurring the image while reducing high-frequency noise. Median filters work differently; though not inherently a frequency domain operation, if conceptualised in this domain, they can be imagined as selectively preserving the median frequency amplitude within a local neighbourhood, which effectively reduces impulse noise without significantly blurring the edges. The Sobel filter, commonly used for edge detection, employs a different approach by computing the gradient of image intensity in the spatial domain to identify edges, which correlates to enhancing high-frequency components when viewed in the frequency domain. In the studies discussed in chapters 2 and 3 of this thesis, we utilised the Butterworth filter (Butterworth, 1930). The Butterworth filter is distinguished by its smooth continuous frequency response that gradually declines past a defined cut-off frequency. This provides a more natural filtering process with minimal

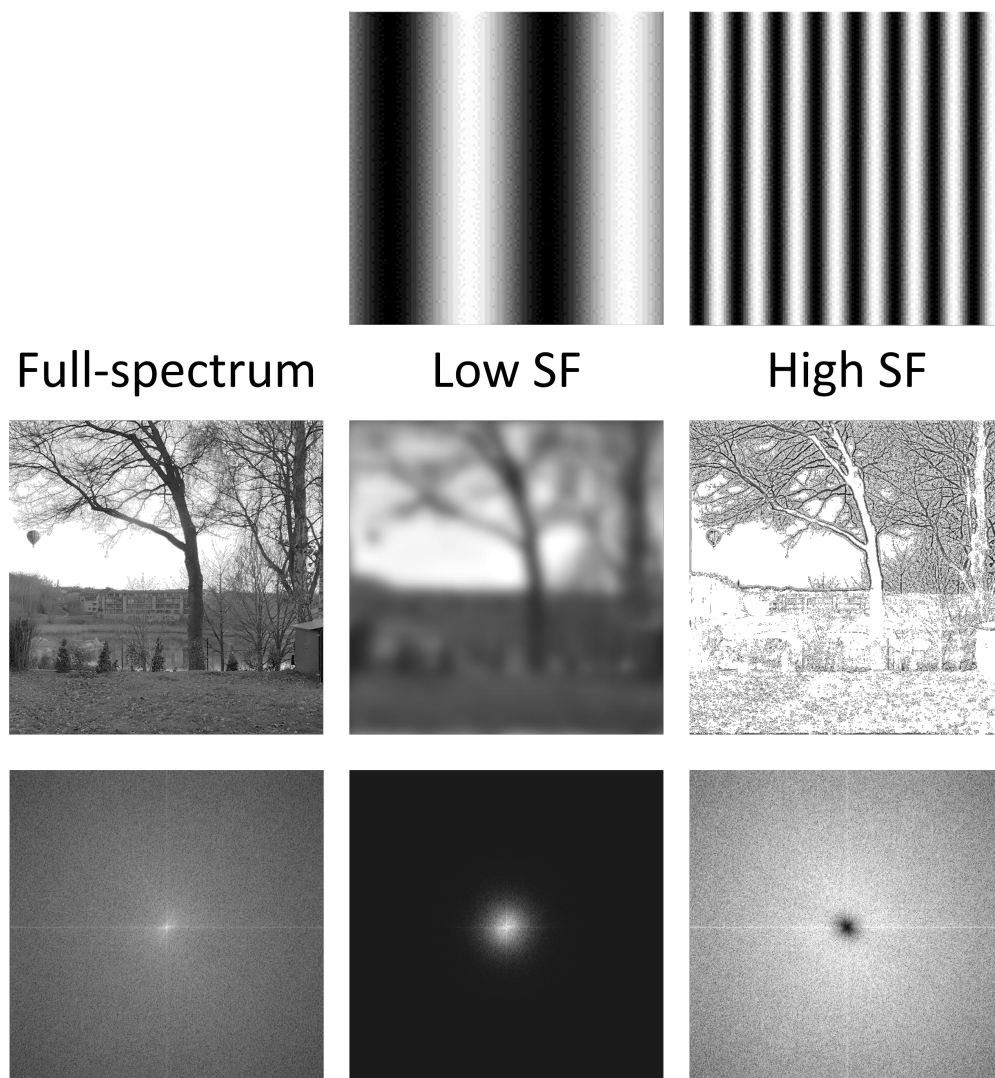


FIGURE 1.1: Spectral analysis of spatial frequencies in greyscale images. The top row illustrates basic, greyscale gradients capturing varying SFs. The image on the left demonstrates a LSF pattern, characterised by wide and smooth gradations, while the image on the right exhibits a HSF pattern, indicated by narrow and closely-spaced gradations. The middle row presents a greyscale depiction of a complex image (an outdoor scene from my garden). From left to right, the first image is full-spectrum, encompassing all SFs. Adjacent to it are images processed through a Fourier Transform and multiplied with Butterworth filter to isolate low and high SFs, respectively. The bottom row translates the middle-row images into the Fourier domain. White signifies high power of the respective SF range and black low power. The full-spectrum image (left) illustrates the prominence of LSF (at the centre of the Fourier) and of cardinal directions (Coppola et al., 1998; Torralba & Oliva, 2003). The Fourier representation of the LSF-filtered image shows the strong LSFs and attenuated HSFs. The representation of the HSF-filtered image depicts amplified HSFs in the Fourier space periphery and a reduction in the central region, indicating a suppression of LSFs from the original image.

ringing artefacts, due to its gentle roll-off characteristics, ensuring a more accurate representation of the image's SF components during filtering.

The filtering process can be divided into three main categories: low-pass, high-pass, and bandpass. Low-pass filters are equipped with a predefined cut-off frequency and have a filter matrix that tapers off the amplitude waves with frequencies higher than this cut-off. This operation lets through the LSF components, while reducing the HSFs, effectively blurring the image and emphasising coarse structures (Figure 1.1, middle row/middle image). Conversely, high-pass filters have a filter matrix that suppresses the LSFs beneath a certain cut-off while preserving HSFs. By doing so, they retain fine details and sharp edges that correspond to the higher frequency waves, which enhances the image's fine structure and edge definition (Figure 1.1, middle row/right most image). Bandpass filters utilise a filter matrix with a specific bandpass defined by a lower and an upper cut-off frequency. Frequencies that fall outside this band are attenuated. Consequently, only SF components within this bandpass are maintained, effectively isolating image details at certain spatial scales.

Information processing in the brain

Hierarchical organisation

In order to understand how our brain interprets visual input, it is crucial to familiarise ourselves with the architecture of the visual system. Traditionally, this system has been perceived as a step-by-step hierarchy, where the initial stages in this hierarchy are tasked with identifying basic features such as oriented edges or bars (Hubel et al., 1978). This fundamental data is then conveyed to advanced visual areas for more complex analysis. Such a bottom-up, or feedforward model of visual processing has been a cornerstone in considerable advancements in this field (Riesenhuber & Poggio, 2000).

From the moment light enters our eyes, it first passes through the cornea, then the lens, and finally projects onto the retina. The retina, situated at the back of the eye, transposes light into neuronal code (transduction), a process essential for processing important features of the visual input (Meister & Berry, 1999). This area of the eye contains two types of photoreceptor cells: rods and cones. The receptive field of these photoreceptors responds to a small patch of the visual field (Hartline, 1938). Rods are widely distributed across the retina, except for the central or foveal region, and are more responsive to variations in light and dark, forms, and movement. The rods are unable to discern colours and are sensitive to LSFs of visual input. Conversely, cones are primarily concentrated in the foveal area and are capable of colour perception and processing visual input with high spatial resolution, although they are less responsive to light intensity. The distribution of cones and rods across the retina is highly skewed, reflecting the specialised functions of the fovea and retinal periphery (Osterberg, 1935). These photoreceptor cells synthesise this light into electrical signals and pass these signals to bipolar cells in the retina, then to the ganglion cells, whose long axons

join to form the optic nerve. The ganglion cells include distinct types: midget cells, which form the parvocellular ("P") pathway, and parasol cells, which form the magnocellular ("M") pathway (Hendry & Yoshioka, 1994; Kaplan & Shapley, 1986; Polyak, 1941). Additionally, there are cells that give rise to the koniocellular ("K") pathway, which conveys information about certain aspects of colour and low spatial resolution patterns. The rods are predominantly connected to the larger, faster-conducting magnocellular cells, which are exceptionally responsive to motion and depth information, but also link with koniocellular and other ganglion cell types that contribute to other visual functions, such as blue-yellow colour contrast. On the other hand, the cones primarily feed into the smaller, slower parvocellular cells that are involved in the perception of colour and fine details, but also engage with the koniocellular pathway and additional cell types (Livingstone & Hubel, 1988; Merigan & Maunsell, 1993). The axons of these cells then travel in the optic nerve to the lateral geniculate nucleus (LGN) of the thalamus, the first major relay and processing centre in the visual pathway. The LGN has corresponding magno-, parvo-, and koniocellular layers that receive input from the respective types of ganglion cells. The receptive fields of neurons in the LGN are similar to those in the retina, maintaining a consistent representation of the visual field (Hubel, 1960). The LGN, in turn, sends these signals to the primary visual cortex (also called V1 or striate cortex) located in the calcarine fissure in the occipital pole. Neurons in V1 are spatially organised according to their receptive field properties (Hubel & Wiesel, 1962). At this stage, the information relayed to V1 is still very basic, such as orientation and direction of movement. V1 neurons are notably selective for specific orientations of luminance contrast, forming the basis of image contour analysis (Hubel & Wiesel, 1962; Tootell et al., 1995). This contour detection ability is enabled by the structure of the neuron's receptive fields, which are larger and elongated with an 'on-centre' area derived from the concentric 'on-centre' input from ganglion cells (Hubel & Wiesel, 1962). The representation of SF in V1, is proposed to be structured into diverse SF selective domains, with imaging and electrophysiological studies identifying consistent clusters of these domains across the cortical surface (DeAngelis et al., 1999; Everson et al., 1998; Hübener et al., 1997; N. P. Issa et al., 2000; Shoham et al., 1997; Tolhurst & Thompson, 1982; Tootell et al., 1981; Xu et al., 2007). This organisation is not uniformly radial from the cortical surface to white matter; rather, a variety of SF domains are arrayed across the cortex, suggesting a more complex mapping than initially believed (Maffei & Fiorentini, 1977). Earlier studies highlighted the presence of low and high SF domains, while later research, accounting for differences in stimuli and imaging postprocessing techniques, showed a broader spectrum of SF preference domains (Everson et al., 1998; N. P. Issa et al., 2000; Xu et al., 2007). The architecture of these SF maps has been debated, with some suggesting a modular structure of alternating patches resembling ocular dominance columns, while others propose a pinwheel organisation akin to orientation columns, or even a combination of target-like patterns with low and high SF domains usually separated by intermediate frequency domains (Everson et al., 1998; N. P. Issa et al., 2000; Shoham

et al., 1997; Xu et al., 2007).

The cortical visual system is characterised by two main processing streams that emerge from the primary visual cortex: the dorsal "where" pathway, associated with spatial awareness and motion processing and the ventral "what" pathway, related to object identification and recognition (see Figure 1.2, and e.g. Milner and Goodale, 1995; Milner and Goodale, 2008; Mishkin et al., 1983; Ungerleider and Haxby, 1994). Contemporary understanding suggests that the delineation between the magnocellular and parvocellular pathways contributing to these streams is not exclusively partitioned; both pathways interact and project to multiple regions across both streams (Goodale & Milner, 1992; Kravitz et al., 2013). The magnocellular pathway, known for its fast response to changes in luminance and high temporal resolution, contributes significantly to the dorsal stream, supporting tasks related to spatial localisation and motion detection. Regions within this stream, such as the middle temporal area (MT or V5) and the medial superior temporal area (MST), are involved in the perception of motion. The MT processes speed and direction, playing a key role in motion coherence perception (Dubner & Zeki, 1971; Newsome & Pare, 1988; Snowden et al., 1992; Zeki, 1974), and MST analyses more complex motion characteristics and spatial orientation (Duffy & Wurtz, 1991; Komatsu & Wurtz, 1988). The intraparietal sulcus (IPS) plays an essential role in integrating spatial information, which is crucial for coordinated actions such as reaching and grasping (Colby et al., 1996; Galletti et al., 1997; Snyder et al., 1997; Taira et al., 1990; Taira et al., 2000). The dorsal stream further extends into the posterior parietal cortex (PPC), implicated in visuospatial attention (Bushnell et al., 1981). The ventral stream, associated with high spatial resolution and colour perception, is predominantly associated with the parvocellular pathway, and implicated in the discrimination of fine visual details and object recognition. However, this pathway too receives input from the magnocellular system, underscoring its role in swiftly processing visual information that facilitates recognition even under rapidly changing conditions. Within the ventral stream, the inferotemporal (IT) cortex contains specialised areas for processing of diverse visual stimuli, each with unique functional roles in processing distinct categories of visual stimuli (Epstein & Kanwisher, 1998; Grill-Spector & Weiner, 2014; Kanwisher et al., 1997; Puce et al., 1996). The parahippocampal place area (PPA) handles processing of scenes and is sensitive to the spatial layout of a scene and helps to differentiate between different types of environments (Epstein & Kanwisher, 1998). Another region, the visual word form area (VWFA), specialises in the perception and recognition of written words, providing the basis for reading and literacy skills (Cohen et al., 2000). Additionally, the fusiform face area (FFA) specialises in processing of faces (e.g. Kanwisher et al., 1997; McCarthy et al., 1997). Beyond these regions, the ventral stream encompasses the lateral occipital complex (LOC), involved in object shape perception (Malach et al., 1995), and the occipitotemporal sulcus, associated with the processing of texture and detailed object features (Ishai et al., 1999). The occipitotemporal area also contains regions sensitive to complex object characteristics, such as the extrastriate body area (EBA) which specialises

in processing body parts and human movement (Downing et al., 2001).

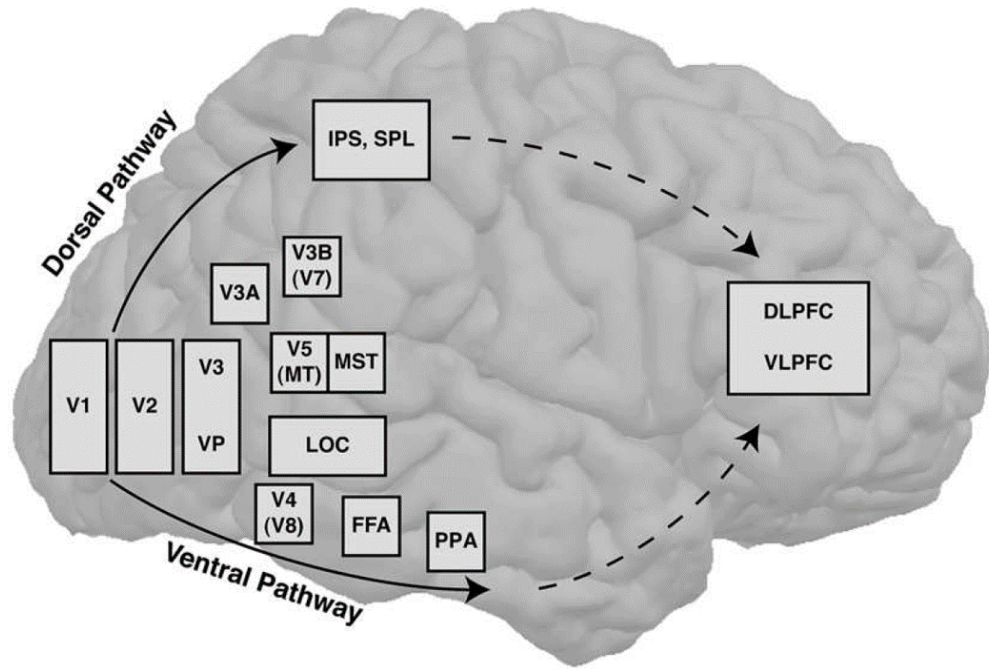


FIGURE 1.2: Depiction of dorsal and ventral cortical pathways; Figure reprinted from Goebel et al. (2004). The architecture of the brain's visual areas, with a dual-pathway organisation consisting of a dorsal (occipitoparietal) and ventral (occipitotemporal) pathway, represented by solid lines. Both pathways are presumed to reach into the prefrontal cortex, as denoted by dashed lines.

Traditionally, researchers believed that visual processing occurred in a feedforward manner (DiCarlo et al., 2012; Hubel & Wiesel, 1968; Riesenhuber & Poggio, 1999). According to this perspective, high-level visual perception is an incremental and unidirectional aggregation of visual information as signals progress from lower to higher regions in the visual hierarchy (Kobatake & Tanaka, 1994; Op De Beeck & Vogels, 2000; Riesenhuber & Poggio, 1999; Rust & DiCarlo, 2010; Serre et al., 2007). This aggregation process allows higher processing stages to focus on more abstract aspects of visual stimuli, detaching from the variable properties of primary visual representation to extract the inherent meaning of the input. Such a mechanism ensures that regions higher within the hierarchy exhibit tolerance to appearance variations due to factors like lighting conditions or occlusion. From this perspective, the fundamental dimensions extracted at the lower levels, such as the V1, are aggregated and become inaccessible at the higher stages of processing (Keyser et al., 2001; Riesenhuber & Poggio, 1999; Serre et al., 2007). Within this theoretical framework, the role of V1 is essential for the initial stages of visual processing, where it is instrumental in forming initial representations of visual input. This viewpoint offers an explanation for the emphasis of current models of perception on higher-order processing (Bruce & Young, 1986; Haxby et al., 2000; Jiang et al., 2006), sometimes at the expense of acknowledging the influence that low-level visual mechanisms may exert on perceptual outcomes.

Recurrent processing

The established model of visual processing, heavily influenced by Marr's pioneering work (Marr, 1982), has been enhanced with detailed anatomical and functional insights. The understanding of cortical connectivity in the visual system has evolved substantially, recognising not just traditional feedforward connections, but also acknowledging the complex web of feedback and lateral connections that underpin visual processing (Markov et al., 2011). For example, there is a myriad of evidence for the existence of anatomical cortical afferents to V1, supported by tracing studies (Burkhalter & Bernardo, 1989; Coogan & Burkhalter, 1990; Rockland & Virga, 1989; Van Essen et al., 1992), animal and human electrophysiology (Bastos et al., 2015; Bullier et al., 1988; Kok et al., 2014; Michalareas et al., 2016), and neuroimaging research (Kok et al., 2012; Muckli et al., 2015). These findings challenge the conventional perspective that complex visual representations are formed solely through a one-way feedforward process. Instead, a comprehensive array of findings argues for a recurrent processing model. According to this alternate view, visual perception is not unidirectional but rather a complex interplay of feedforward, feedback, and lateral interactions across various processing levels (Ahissar & Hochstein, 2004; Bullier et al., 1988; Edelman & Gally, 2013; Lamme et al., 2002; Mumford, 1992).

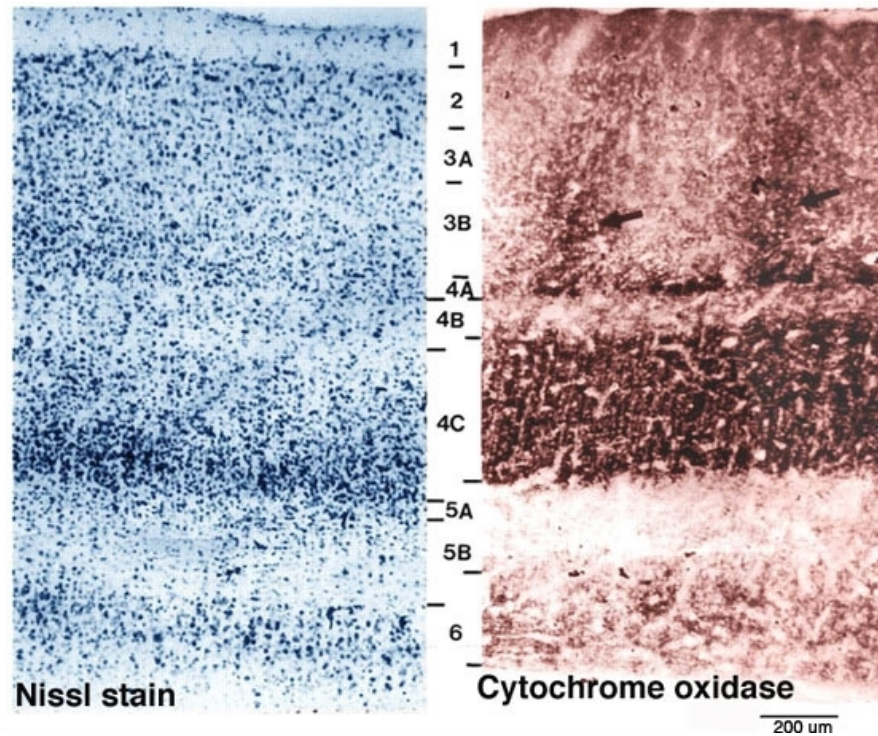


FIGURE 1.3: Laminar structure of V1 in a macaque monkey. The figure is adapted from images adapted from Kolb et al. (1995). The left panel displays a Nissl stain while the right panel highlights cytochrome oxidase labelling, delineating the complex layering of V1. The visual cortex is stratified into six main cortical layers according to Brodmann's classification, with layer 1 at the top, closest to the surface, and layer 6 at the bottom, bordering the white matter (Douglas & Martin, 2004; Hawken et al., 2020; Self et al., 2013; T. Wang et al., 2020). The most superficial layer, layer 1, is notably sparse in neurons and is primarily a nexus for dendrites and axons. In contrast, layers 2 through 6 houses around 20% inhibitory interneurons pivotal to the internal circuitry of V1 without projecting externally. The supergranular layers, layers 2 and 3, are rich in excitatory neurons that project their axons to higher-order visual areas such as V2, V3, V4, and MT. Layer 4, distinguished by its granular appearance, is further divided into sublayers 4A, 4B, 4C α , and 4C β , with sublayers 4C α and 4C β being the principal targets of input from the LGN (Blasdel & Lund, 1983; Callaway, 1998; Hubel & Wiesel, 1972). This input maintains the segregation of the M- and P-pathways, with the M-pathway synapsing in 4C α and the P-pathway in 4C β . The infragranular layers, layers 5 and 6, contain projection neurons that link back to the LGN, forming a feedback loop. Notably, layer 4 is the primary beneficiary of axons from the LGN, although these axons can also reach layers 1, 3, and 6. The organisation within V1 reflects similarities with the retina and LGN.

While V1's initial tuning is guided by the LGN input, the recurrent processing framework emphasises that the actual refinement of visual processing in V1 is heavily influenced by a dynamic balance of intracortical excitatory and inhibitory mechanisms (e.g., Benevento et al., 1972; Ben-Yishai et al., 1995; Carandini and Ringach, 1997; Somers et al., 1995; Worgotter and Koch, 1991). V1, with its distinct laminar organisation (Figure 1.3; Douglas and Martin, 2004; Hawken et al., 2020; Self et al., 2013; T. Wang et al., 2020), has been known to receive feedforward inputs primarily targeting layer 4C from the LGN (Blasdel & Lund, 1983; Callaway, 1998; Hubel & Wiesel, 1972). These inputs carry the initial visual cues that are foundational for the perception of intensity and basic features. However, this is merely the starting point for visual information processing. Recurrent processing refines these initial visual cues by iterative exchanges within V1 and between V1 and other visual areas. These exchanges occur through a complex web of feedback and lateral connections, both excitatory and inhibitory, involving multiple layers such as layers 2/3, 4B, 5, and 6. The role of recurrent processing is multifaceted, enhancing the accuracy of visual representations, aiding in perceptual inference, and providing stability in visual perception despite ambiguous or incomplete visual input. Feedback connections from higher visual areas such as V2 and V4, and those associated with attention and memory systems, are crucial for integrating global contextual information. They have a strong presence in V1's superficial (2/3) and deep layers (5/6) and can be excitatory, to enhancing visual representations, or inhibitory, to filter out irrelevant stimuli, thus maintaining focus on pertinent aspects of the visual field (Anderson & Martin, 2009; Gilbert & Wiesel, 1983; Maniglia et al., 2019; Rockland & Pandya, 1979; Rockland & Virga, 1989). Feedback pathways, therefore, serve not only to enrich the visual representation with context and anticipation but also to conduct error correction by aligning the visual perception with expectations and memory-based predictions (Bastos et al., 2012; T. S. Lee & Mumford, 2003; Lochmann & Deneve, 2011; Murray et al., 2002; Rao & Ballard, 1999; Schwiedrzik & Freiwald, 2017). In this 'predictive coding' framework, the feedback signal is predictive while the feed-forward input is interpreted as a prediction error (Keller & Mrsic-Flogel, 2018); feedback from higher-level visual regions is thought to suppress redundant activity in early processing stages, this 'explaining away' of activity simplifying the encoding of the visual stimulus (Harrison et al., 2007; Murray & Wojciulik, 2004). With respect to models of figure-ground segregation, feedback is posited to selectively amplify V1 responses to segments of the visual field that are identified as 'figure' as opposed to 'ground' (Corthout & Supèr, 2004; Hupé et al., 1998; Rossi et al., 2001). Furthermore, attention-driven models indicate that feedback can fine-tune or anticipate V1 activity in response to visual stimuli (Ito & Gilbert, 1999; Lamme et al., 1998; Luck et al., 1997; Ress & Heeger, 2003). Others contend that feedback serves to refine perceptual representations (Angelucci & Bressloff, 2006; Angelucci et al., 2017; Bair et al., 2003; Bullier et al., 1996; Hupé et al., 1998; Jehee et al., 2007) or to enhance representations that become part of our conscious experience (Supèr et al., 2001).

Lateral (or horizontal) connections within V1, particularly prominent in the upper layers 2/3 but also extending throughout layers 4A, 4C and 6, foster the integration of features and aid in the modulation of visual processing (Gilbert & Wiesel, 1983; Rockland & Lund, 1983). These connections are essential in achieving iso-feature enhancement (excitatory signals) and lateral inhibition (Gilbert et al., 1990; McGuire et al., 1991; Polat & Sagi, 1993; Rockland & Lund, 1982), processes that underpin our ability to perceive coherent visual scenes and distinguish contrasting elements. Moreover, the impact of feedback pathways expands with distance from V1; signals from further areas encapsulate larger visual fields, which convey more extensive contextual information to a given V1 column than horizontal connections do. The hierarchical organisation of the visual system allows for a progressive and nuanced interpretation of visual stimuli, from the simple to the complex, shaped by feedback that modulates and anticipates the feedforward flow.

In support of the recurrent processing model, longer-latency responses observed in V1 neurons reflect top-down attentional influences (Hupé et al., 1998; Ito & Gilbert, 1999; Motter, 1993; Roelfsema et al., 1998) and higher-order perceptual contexts, such as the perception of 'figure-ground' (Lamme, 1995; Zipser et al., 1996), sensitivity to subjective contours (T. S. Lee & Nguyen, 2001), and discernment of shapes from shading (T. S. Lee et al., 2002). In line with these studies, (Lamme, 1995) investigated the modulation of a cell's response by extrinsic contextual information in a figure-ground experiment. Their experimental design featured a textured figure superimposed on a background of either the same texture, diminishing any distinction between figure and background, or a differently oriented background, emphasising figure-ground distinction. Though the contents of the neuron's receptive field were not sufficient for the neuron to differentiate between the figure or the background, an interesting pattern emerged. Initially, the cell responded similarly to both conditions. However, after approximately 100ms, an enhanced response was observed for the figure set against a different background. This phenomenon of figural response enhancement, emanating from outside the receptive field and accompanied by a temporal delay, serves as a compelling testament to the role of recurrent processing in visual perception. Another illustration of the role of recurrent processes in visual perception can be seen in a study Kok and de Lange (2014), where researchers explored the spatial profile of neural activity in V1 during the perception of illusory shapes. They studied this with the Kanizsa triangle (Kanizsa, 1976), where three "Pac-Man" shapes incite the illusion of a triangle overlaying three disks (Figure 1.4). Despite the absence of explicit contours or an increased brightness, people perceive a clear triangle. Kok and de Lange (2014) demonstrated that the perception of this illusory triangle enhanced neural activity in V1 regions receptive to the shape but devoid of bottom-up/feedforward input, and it suppressed activity in regions receiving bottom-up input (Friston, 2005; Rao & Ballard, 1999). This shows that the perception of shape modulates responses in early regions of the visual hierarchy, implying the existence of recurrent processes.

To study the timing of recurrent processes, researchers adopt methods that

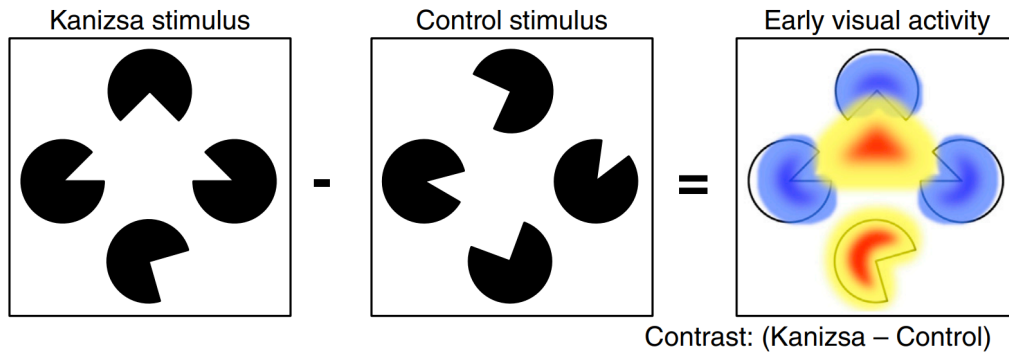


FIGURE 1.4: Kanizsa triangle and fMRI responses; figure adapted from Bartels (2014). On the left, the Kanizsa triangle is presented, where three of the four "Pac-Man" shapes are arranged in such a way that an observer perceives an illusory triangle. In contrast, the control configurations (centre) show the shapes misaligned, eliminating the perception of any illusory figure. The right image illustrates the reconstruction of fMRI responses in the V1 area, specifically triggered by the illusory triangle configuration in line with findings from Kok and de Lange (2014). The V1 regions corresponding to the illusory figure experience enhanced neural activity when the triangle is perceived, compared to configurations where the shapes do not form any discernible figure.

temporarily disrupt cortical processing. It has been found that there is engagement of recurrent feedback circuits within the first 80-150ms of visual processing. One significant source of evidence is derived from experiments using transcranial magnetic stimulation (TMS), a technique that can temporarily inhibit a targeted brain area's responsiveness. For instance, a study conducted by Koivisto et al. (2011) leveraged TMS, guided by fMRI localisation, to selectively inhibit V1/V2 while subjects categorised images based on whether they contained an animal or not. The researchers discovered that administering TMS to V1/V2 with stimulus onset asynchronies (SOAs) of 90-210ms hindered both categorisation performance and the subjective perception of stimuli. Camprodon et al. (2010) reported similar findings when they applied TMS over V1 only. They found two periods of impairment, with SOAs at around 100 and 220ms. This aligns with earlier work from Corthout et al. (1999), who also identified an early window of activity at around 100ms during which TMS application over V1 impaired letter recognition (Corthout et al., 1998, 1999). Taken together, these experiments show that early visual areas' disruption around 100ms post-stimulus presentation can impair visual recognition. Interestingly, this time window occurs after the earliest contributions of IT neurons, suggesting the possibility that the impairment results from disrupting feedback to lower-level areas, thereby influencing the quality of object representations. Moreover, several studies identified a second, later time window around 200ms, during which TMS also impaired recognition. This latter window coincides with the latency of spatial attention-mediated processing (see e.g. Buffalo et al., 2010; Martinez et al., 2001; Mehta et al., 2000a, 2000b; Noesselt et al., 2002).

Degraded stimuli

The need for recurrent processing in the visual system becomes even more apparent when stimuli are ambiguous, incomplete, or otherwise degraded. Various studies have illustrated that the responses in object-selective areas, such as those in the IT cortex, are lower when a stimulus is degraded (Kovács et al., 1995; Nielsen et al., 2006; Sclar et al., 1990; Williford & Maunsell, 2006). Recurrent processing over time has been found to not only restore but nearly normalise these responses. This recovery to near-original levels is accomplished through top-down reinforcement, preserving the selectivity of the responses (Akrami et al., 2009; Daelli et al., 2010). Additionally, object-selective responses in the IT cortex are strong when stimuli are fully visible, but take noticeably longer to develop than when the stimuli are obscured an average increase of around 50ms (Kovács et al., 1995; Nielsen et al., 2006).

This dynamic is mirrored in V1, where reduced feedback from higher-order visual areas correlates with increased neural activity, particularly in the context of scrambled stimuli (Goffaux et al., 2016; Lerner et al., 2001). This phenomenon was also observed by Freud et al. (2017), who measured functional magnetic resonance imaging (fMRI) responses to object images with varying degrees of visual scrambling. The researchers quantified shape sensitivity as slope values, i.e. responses as a function of scrambling level. The analysis of these beta weights revealed that the most posterior voxels displayed a low sensitivity to object shape, indicating high responses in early visual regions in response to scrambled images. These findings align with the predictive coding framework, supported by work from Murray et al. (2002), this framework suggests that coherent object shapes elicit relatively low responses in V1 and heightened activity in the LOC. This pattern suggests a model where higher-order visual areas modulate activity in the primary visual cortex, potentially inhibiting irrelevant responses through feedback mechanisms.

The importance of feedback mechanisms is also highlighted by experiments that manipulate activity in higher-level areas. For example, Hupé et al. (1998) used reversible cooling to temporarily disable V5/MT. Recordings from V1 to V3 showed that responses to moving bar stimuli were significantly reduced (fewer spikes observed per second) when V5/MT was deactivated compared to control conditions in which it remained active. This suppression of lower-level responses was most pronounced under low salience conditions, such as when the bar stimulus had a very low contrast. These results suggest that when higher-level motion selective visual areas are active, they deliver additional excitatory input to lower levels.

Complementary to these findings, computational modelling studies provide support to the role of enhanced recurrent processing in the visual system in response to stimulus degradation. Initial visual responses to a degraded stimulus may initially mirror the distorted information picked up by the fovea. With time, recurrent processing amplifies the significant features of the stimulus and suppresses the irrelevant ones, thereby facilitating the categorisation of the altered visual input (O'Reilly et al., 2012; O'Reilly et al., 2013; Rajaei et al., 2019;

Wyatte et al., 2012). Furthermore, a behavioural study in humans conducted by Wyatte et al. (2012) shed light on the importance of recurrent processing in visual object recognition. In this study, stimuli were impaired through occlusion and contrast degradation, and they introduced disruption to recurrent processing using backward masking (Enns & Di Lollo, 2000; Lamme & Roelfsema, 2000). The researchers found that when a mask immediately followed a severely degraded stimulus thus interfering with recurrent processing the ability to categorise the stimulus was significantly hampered. However, when a clear stimulus was presented, categorisation performance remained consistent.

In short, the collected findings underscore the essential role of recurrent processing in visual perception, providing a dynamic mechanism to strengthen relevant and suppress irrelevant visual signals, particularly when faced with ambiguous or degraded stimuli, thereby ensuring consistency and accuracy in object recognition.

Coarse-to-fine processing

Various researchers propose that recurrent processing exploits the alignment of SFs to facilitate efficient processing of visual information (Bar, 2004; Bullier, 2001; Hegdé, 2008; Kreiman & Serre, 2020). Within this framework, the rapid feedforward transmission of a scene or object's coarse structure –represented by LSFs and mainly conveyed via the fast, magnocellular pathway –ignites visual processing. During the early stages of processing, spanning only a few tens of milliseconds, an initial examination generates a coarse representation of the visual input in advanced dorsal and frontal visual regions, with latencies of approximately 90ms and 100ms, respectively (Bullier, 2001; and see Figure 1.5). The LSF representation enables early predictions about the objects and context responsible for such visual input (Bar, 2004; T. S. Lee, 2015). At a latency of approximately 110ms, the formed predictions activate corresponding visual representations in object processing regions in the IT cortex. To facilitate recognition of an object, the bottom-up processes are guided by feedback signals to concentrate on a specific set of the most probable object representations. Initial high-level coarse representations, shaped from LSF information, are subsequently sent back to V1 and V2. Such feedback serves to guide the processing of fine-grained cues held within HSF input, delivered about 20ms later through the slower and more sustained parvocellular pathway (Baseler & Sutter, 1997; Nowak et al., 1995; Schmolesky et al., 1998). In this framework, the V1 (and V2) is postulated to function as an 'active blackboard' for the coarse-to-fine integration of visual information (Bullier, 2001).

One condition for coarse-to-fine processing is that LSF processing takes precedence over HSF processing. Empirical evidence supporting this sequence has been observed with simple stimuli, such as dots, lines and gratings, which evoke faster responses when they are of lower SFs (e.g., Breitmeyer, 1975; Hughes et al., 1996; Jones and Keck, 1978; Mihaylova et al., 1999; Musselwhite and Jeffreys, 1985; Parker, 1980; Parker and Dutch, 1987; Watt, 1987). For example, Parker and

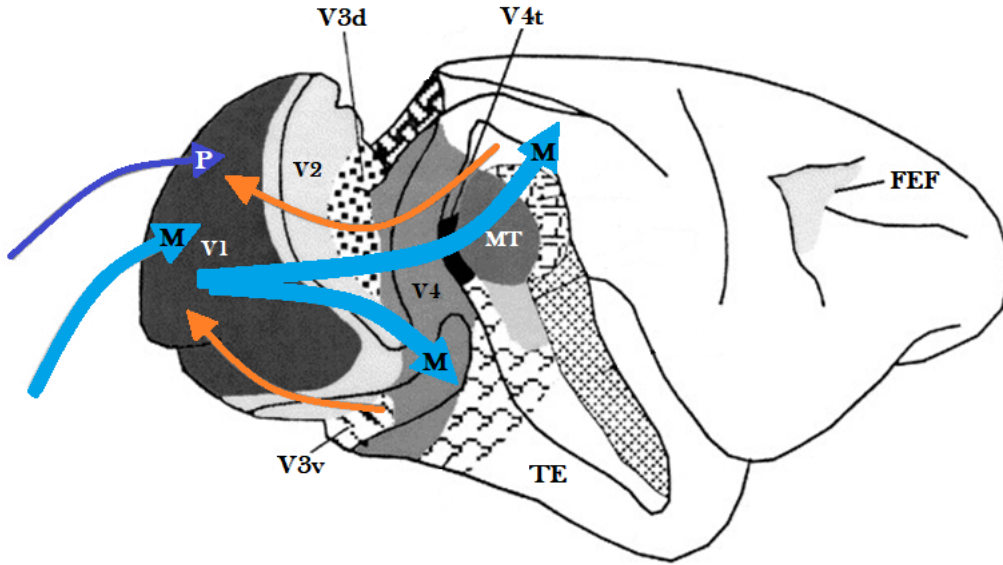


FIGURE 1.5: Proposed model of integrated visual processing; figure adapted from Bullier (2001). The depicted big light blue arrows symbolise the neural activity transmitted through the M-pathway, initiating the activation of visual areas in a sequential manner. Feedback connections originating from the parietal and temporal cortex are hypothesised to convey signals that are retro injected into lower-order areas (thin orange arrows), synchronising with the arrival of the narrower dark arrow representing the P wave of activity.

Salzen (1977) measured electrophysiological responses on occipital sites in relation to the presentation of sine wave gratings at different SFs. They observed that as the SF of the grating increased, the latency of the peak activity generated by the stimulus presentation also increased. Additional support of the coarse-to-fine precedence can be found for complex stimuli in higher-level brain regions, such as in the FFA in response to faces or PPA for natural scenes (e.g., Goffaux et al., 2011; Halit et al., 2006; Kauffmann, Chauvin, Guyader, and Peyrin, 2015; Kauffmann, Ramanoël, et al., 2015; Lu et al., 2018; Musel et al., 2014; Parker et al., 1992, 1997; Peyrin et al., 2006; Schyns and Oliva, 1994; Vlamings et al., 2009). For instance, LSF content in faces drives early electrophysiological differences between faces and other visual categories as well as behavioural markers of face-specialised processing, while HSF is optimal for analysing feature details (e.g., Goffaux et al., 2003, 2005; Goffaux and Rossion, 2006; Quek et al., 2018; Vlamings et al., 2009).

While the temporal precedence of LSF over HSF is well documented, evidence for the integrative aspect of the coarse-to-fine model is less abundant. Coarse-to-fine processing entails not only the rapid processing of coarse image information compared to finer details, but it also requires that this initial processing informs and influences the subsequent processing of HSF information. The benefit of coarse-to-fine integration should be a reduction in both processing time and metabolic cost of visual processing by conserving resources that

would otherwise be dedicated to processing redundant detailed HSF information. To investigate coarse-to-fine integration experimentally, it is necessary to mimic several key characteristics of natural visual input. Firstly, natural images are full-spectrum/broad-band, encompassing both low and high SFs. Secondly, the different ranges of SFs that make up an image are typically highly correlated in space. Without such correlation, coarse-to-fine integration would not be computationally beneficial. Past studies have often examined coarse-to-fine processing by comparing neural responses to stimuli filtered for specific SF ranges. While these studies provide solid evidence for the precedence of coarse over fine detail processing, they fall short in addressing coarse-to-fine integration because the SF ranges were presented separately in space, time, or both (e.g. De Cesarei and Codispoti, 2013; Goffaux et al., 2011; Musel et al., 2014). For example, some studies have employed hybrid stimuli (e.g. Kauffmann et al., 2017; Rotshstein et al., 2007; Schyns and Oliva, 1994), which consist of images created from two superimposed scenes of different categories and containing different SF bands (such as a city scene in LSF combined with a highway scene in HSF). This approach neglects the principle that LSF and HSF signals should be spatially correlated. These studies departed from the principles of naturalistic visual input, which is essential for efficacy and functional relevance of coarse-to-fine integration. To address SF integration of naturalistic stimuli Petras and colleagues conducted two complementary studies employing electroencephalography (EEG; Petras et al., 2019) and magnetoencephalography (MEG; Petras et al., 2021). In these studies, subjects were presented with broad-band images of faces, as well as phase-scrambled versions filtered to isolate either LSFs or HSFs. Classifiers were trained using the neural EEG/MEG patterns evoked by filtered scrambled stimuli, and subsequently applied to predict de categorisation of both broad-band (intact) and scrambled trials. The findings from these studies show that when broad-band face stimuli are presented, LSFs contribute to the neural representation before HSFs, implying that LSF information may play a guiding role in the processing of HSF details. Moreover, the presence of LSF information in broad-band images correlates with a reduction in the dominance of HSF representations (Petras et al., 2019) in both early and higher-level visual areas (Petras et al., 2021). The decrease in gamma-band activity they found in the early visual cortex suggests that LSF information can indeed guide the processing of HSF details by utilising the cross-SF redundancies inherent in natural images, thereby enhancing processing efficiency. However, due to the limited spatial resolution of EEG and MEG, the involvement of V1 in particular in this coarse-to-fine integration could not be definitively determined.

Backward masking

Backward masking to study recurrent processing

One method that shows promise in studying recurrent processing is the technique known as backward masking. This approach involves the rapid sequential presentation of two stimuli: an initial "target" stimulus followed by a "mask" stimulus. The mask can be presented at varying SOAs, where SOA denotes the time interval between the start of the first stimulus and the start of the mask. Varying the masking SOA can alter the visibility of the target stimulus. For example, with a sufficiently short SOA, the mask can substantially impair target recognition, and under certain circumstances, may even prevent its perceptual awareness (Bridgeman, 1980; Dehaene et al., 2001; Haynes et al., 2005; Kovács et al., 1995; Lamme & Roelfsema, 2000; Macknik & Livingstone, 1998; Rolls & Tovee, 1994; Rolls et al., 1999; Thompson & Schall, 1999). Research into the temporal dynamics of visual processing using backward masking demonstrated that, under short SOA conditions, the neural response in the non-human primate temporal cortex is transient. For instance, with SOAs as brief as 20ms, face-selective neurons exhibit activity for merely 20 to 30ms, indicating that the mask disrupts the processing of the target stimulus (Rolls & Tovee, 1994). In line with these findings, Kovács et al. (1995) found an improvement in face recognition accuracy as the SOA increases from 20 to 40ms in human subjects, hinting at the delicate balance of timing in visual recognition tasks.

The neural underpinnings of backward masking have spurred debate and multiple theories. Early investigations of backward masking effects centred around feedforward mechanisms (Breitmeyer & Ganz, 1976), such as low-level lateral inhibition, which does not fully account for scenarios where the stimulus and mask fully overlap spatially (Francis, 1997; Macknik & Livingstone, 1998). Alternatively, it has been posited that the disruption caused by masking arises from inter-channel inhibition in the early visual system (Breitmeyer & Ganz, 1976). This view is based on the characteristics of the M-pathway –sensitive to transient stimulation such as rapid motion or sudden onset –and the P-pathway, which favours slow-moving or stationary stimuli (Breitmeyer, 1973; Keesey, 1972; Kulikowski & Tolhurst, 1973; Pantle, 1970; Tolhurst, 1973). The rapid onset of the mask, processed via the M-pathway, would inhibit the sustained cell activity (the stationary target) through transient cell activation. However, neurophysiological evidence revealed early transient responses to masked stimuli that went undetected in higher visual areas like the IT cortex, as well as in the visual neurons of the frontal eye field (Kovács et al., 1995; Rolls & Tovee, 1994; Thompson & Schall, 1999). This has catalysed a re-evaluation of the mechanics behind backward masking, with current theories emphasising the significance of recurrent processing (Figure 1.6; Enns and Di Lollo, 2000; Lamme et al., 2002; Wyatte et al., 2012). This theory proposes that while backward masking does not typically affect the initial feedforward processing of the target stimulus, it can disrupt the subsequent recurrent interactions (Bridgeman, 1980; Kovács et al., 1995; Macknik & Livingstone, 1998; Rolls & Tovee, 1994; Rolls et al., 1999; Thompson & Schall,

1999). For example, Fahrenfort et al. (2007) illustrated the impact of masking on feedback processes through a figure-ground segmentation paradigm analogous to the one used by Lamme (1995). They used EEG to measure responses of the human visual cortex during the perception of a textured stimulus. Their findings indicated that under masking conditions, the EEG signatures associated with feedback processing (i.e. sustained responses present in unmasked conditions) were notably absent, even though the feed-forward responses remained detectable regardless of the subjects' performance in recognition tasks. These insights led to the conclusion that masking's efficacy stems from its ability to interrupt recurrent processing, thereby undermining the neural underpinnings of figure-ground segmentation. Moreover, tasks presumed to require minimal or no top-down processing (i.e. less recurrent processing), such as detection tasks, have been found to be less vulnerable to masking effects compared to more demanding tasks such as identification (Grill-Spector & Kanwisher, 2005). Further insights come from a study by Wyatte et al. (2012), in which the authors manipulated visual object stimuli through occlusion and contrast degradation, using backward masking to modulate the availability of recurrent processing mechanisms. When clear stimuli were masked with a long latency (100ms SOA), recognition performance was minimally affected. However, when stimuli were heavily occluded or of low contrast (i.e. low signal-to-noise ratio; SNR), the masking effect was significantly greater, implying that recurrent processing was critical for discerning object identity under conditions of ambiguity or degradation. These observations have been supported by computational simulations that included recurrent feedback between hierarchically adjacent layers (O'Reilly et al., 2013). The studies above underscore the pivotal role of recurrent processing, particularly when the task requires the integration of visual information across different processing stages. The interruption of this feedback loop by masking leads to a disconnect in the visual information processed at various cortical levels, effectively hindering the brain's ability to construct coherent visual perception (Boehler et al., 2008; Fahrenfort et al., 2007; Lamme et al., 2002).

Regarding the temporal dynamics of backward masking, we can consider a classing experiment by Bachmann and Allik (1976). This experiment involved presenting subjects with two geometric figures of equal area –serving as the target and the mask –each displayed for a mere 10ms. Their investigation into the subjects' ability to recognise the target across varying SOAs showed interesting patterns. At an SOA of 0ms, where both stimuli appeared simultaneously, the figures were often perceived as a singular, composite object. As the SOA extended to the range of 40-80ms, the visibility of the target diminished, with identification accuracy to near chance levels. This suggests a critical window where recurrent processing is instrumental for the target's identification. It was only at an SOA of approximately 150ms that the target began to be perceived as a stable, independent entity, with consistent identification requiring an SOA of around 250ms-marking the completion of the associated cognitive processes. In line with these findings, Fahrenfort et al. (2007) delineated distinct stages within the masking timeline. An initial phase, occurring before 110ms, involved bilateral

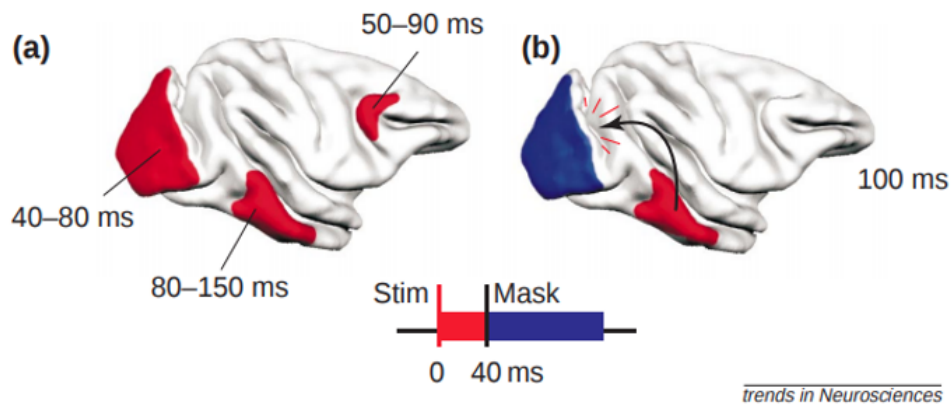


FIGURE 1.6: *Disruption of recurrent interactions by masking stimuli; figure reprinted from Lamme and Roelfsema (2000). The neural response is elicited by a 40ms stimulus presentation (depicted by the red bar), followed by the introduction of a mask stimulus (depicted by the blue bar). Psychophysically, this typically results in a significant reduction in the visibility of the red stimulus. a) The latencies of response intervals associated with the 40ms stimulus (excluding the mask) are depicted for monkey V1, frontal eye field, and IT regions, represented in red. The initial feedforward processing pathway remains relatively unaffected by the masking stimulus. (b) The masking effect may arise due to a mismatch between recurrent and feedforward information: at 100ms, the temporal cortex regions become activated by the stimulus (red) and send feedback signals to lower areas. However, at this stage, V1 is activated by the mask stimulus (blue), disrupting the congruence between the feedback and low-level activity.*

anterior occipitotemporal neural activation, indicative of feedforward processing towards the extrastriate visual areas and further along the visual pathway (Foxe & Simpson, 2002). Subsequently, a phase characterised by a prominent posterior occipital activation emerged after 110ms, signalling re-entrant processing directed towards the early visual areas, potentially involving V1. This was followed by a recurrent bilateral neural activation within the 200-300 millisecond timeframe, likely corresponding to processing in extrastriate regions or beyond. These phases collectively suggest that backward masking can effectively influence visual processing for SOAs extending up to 300ms.

Masking coarse-to-fine

Backward masking also provides the opportunity to explore coarse-to-fine processing. For example, a study by Goffaux et al. (2011) utilised backward masking of faces combined with fMRI to examine temporal variations in the processing of LSF and HSF faces in face-specialised regions. The experiment design comprised of presenting either LSF-filtered faces followed by an LSF mask or HSF faces succeeded by an HSF mask, with the masks shown at varied SOAs between 75 and 300ms. Results pointed towards the modulation of responses to HSF and LSF conditions as a function of SOA in face selective brain regions. Early SOAs demonstrated a higher selectivity for LSFs, while HSF selectivity became more pronounced at later SOAs, supporting coarse-to-fine precedence in the high-level visual cortex. Moreover, backward masking is not limited to studying the coarse-to-fine responses to LSF and HSF in isolation. It also enables the exploration of how coarse-to-fine integration operates with full-spectrum stimuli. An illustrative example of masking separate spatial scales of full-spectrum stimuli is a behavioural categorisation experiment by Bacon-Macé et al. (2005). In this study, subjects were shown full-spectrum images from different categories followed by noise masks of varying spatial resolutions. The task involved identifying whether the displayed image represented an animal. Interestingly, the mask with the finest scale proved to be highly disruptive at early SOAs but less so at later SOAs. In contrast, the mask with the coarsest scale had minimal disruptive effects across all SOAs. This observation highlights the promise of using backward masking to study the intricacies of coarse-to-fine integration, providing insights into how different SF scales are selectively processed over time.

Limitations of previous studies

Existing studies have predominantly used relatively simple stimuli, such as dots and gratings, to explore the involvement of early visual areas (i.e. V1, V2) in recurrent processing. For example, the figure-ground study by Lamme (1995) and Fahrenfort et al. (2007) used grated textures or the simple Kanizsa figures as used in the study by Kok and de Lange (2014). These studies provide crucial insights into the temporal dynamics of visual processing in V1/V2. However, such research has primarily employed basic visual inputs which may not fully represent the complexities of real-world visual environments.

Faces as complex stimuli in research

The limitation of simple stimuli is particularly significant given the complexity and heterogeneity of visual information we encounter daily, with faces being one of the most salient and important types of visual stimuli in our environment. They represent some of the most complex and information-rich stimuli we regularly encounter, serving as vital conduits for personal identity and emotional expression. The symmetrical and standardised nature of faces allows for controlled manipulations in experimental settings, while the human brain's specialised areas for face processing provide a window into both localised and networked neural responses. The extensive body of existing face perception research offers a solid foundation for research.

Traditional models of face processing, rooted in hierarchical, sequential information processing (e.g., Halgren et al., 1999; Haxby et al., 2000), suggest that the interpretation of faces is rooted in a core system, composed of the OFA, FFA, and posterior superior temporal sulcus (pSTS) (e.g., Fairhall and Ishai, 2007; Haxby et al., 2000; Ishai, 2008; Pitcher et al., 2007; Rotshtein et al., 2007). The OFA is thought to undertake the initial parsing of facial features, which are then synthesised in the FFA to form invariant representations, such as identity and gender. Concurrently, the OFA is believed to communicate with the pSTS, which processes dynamic facial aspects vital for social communication. This interconnectivity suggests that a bidirectional flow of information, facilitated by feedback connections, is intrinsic to face processing. This system is complemented by an extended network encompassing regions that, while not exclusively visual, are integrated with the core areas and extract various types of information from faces. These include the intraparietal sulcus, which modulates attention according to gaze direction, the auditory cortex, which assists in speech perception, and limbic structures such as the amygdala, which process emotional cues. Additionally, an anterior temporal region is implicated in processing semantic information about familiar faces, linking to the FFA.

More recent findings have cast doubt on the traditional hierarchical model of face processing, challenged by cases where individuals with damage to the OFA still show active responsiveness in the FFA and the pSTS. Such instances suggest that face recognition can remain intact even if the initial processing stage, as posited by the traditional model, is compromised (Rossion, Caldara, et al., 2003; Steeves et al., 2006). This resilience of face recognition is further substantiated by observations of patients with lesions in the right OFA and FFA who continue to exhibit normal face selectivity in the right pSTS, indicating the presence of alternative pathways for face information within the visual network (Dalrymple et al., 2011; Yang et al., 2016). Moreover, a particular case study reports a patient lacking a discernible OFA in either hemisphere. Remarkably, this patient maintains bilateral activity in both FFAs and the pSTS, reinforcing the notion that early visual cortex connects directly with multiple face-selective regions (Steeves et al., 2006). A re-evaluation of the face processing sequence has been proposed, suggesting that initial processing may occur in the FFA, which then informs the OFA—effectively reversing the traditional OFA/FFA hierarchy (Rossion et al., 2014;

Rossion, Caldara, et al., 2003). This hypothesis gains support from a study with a patient whose right FFA was damaged. Despite an intact inferior occipital gyrus, the expected site of the right OFA, this patient did not exhibit typical OFA activity, suggesting an FFA-first processing route (Delvenne et al., 2004; Rossion, 2009). Investigations into the connectivity between face-selective areas have further elucidated the complex network underlying face processing. For instance, Gschwind et al. (2012) conducted a study using diffusion tensor imaging (DTI) to track the fibres connecting these regions. Their findings indicate links between the OFA and FFA but not between these areas and the STS. Moreover, while direct connections from the early visual areas to the FFA were identified, these were less pronounced than connections to the OFA (Gschwind et al., 2012; Pyles et al., 2013). Functional connectivity research, examining the correlation between activation patterns of these regions, reveals a similar theme of selective interconnectivity (Avidan et al., 2014; Davies-Thompson & Andrews, 2012; Fairhall & Ishai, 2007; Grimaldi et al., 2016; Moeller et al., 2008). Collectively, these studies underscore the nuanced interplay of feedforward and feedback processes in face processing, challenging the notion of a strictly hierarchical feedforward system. The OFA's involvement in associating faces with identity-related semantic content or integrating identity with expression exemplifies this dynamic processing. Electrophysiological evidence suggests that predictive feedback mechanisms operate inherently during face processing, offering a computational perspective on these findings (Eick et al., 2020; E. B. Issa et al., 2018; Kadosh et al., 2011). These studies show that face processing necessitates complex interaction of both feedforward and feedback processes for accurate interpretation instead of merely a hierarchical feedforward model (e.g. Fan et al., 2020; Rossion, Caldara, et al., 2003; Sorger et al., 2007). This complexity is reflected in research that spotlights the role of the OFA not just in feature detection but also in the more complex tasks of linking faces to their associated semantic identities and weaving together aspects of identity with emotional expressions (Kadosh et al., 2011). Furthermore, electrophysiological findings have highlighted the presence of predictive feedback during face processing, a mechanism that appears to operate automatically, underscoring a more integrated and interactive approach to understanding how we interpret faces (E. B. Issa et al., 2018). These insights collectively suggest that face perception is shaped by a rich tapestry of interconnected processes that balance the extraction of detailed facial information with the overarching context provided by past experiences and expectations.

Face processing in early visual areas

To achieve a well-rounded understanding of human vision, it becomes imperative to comprehend how recurrent processing operates when faced with complex stimuli, such as faces. Moreover, the exploration of the role of V1/V2 in the recurrent processing of complex stimuli, such as faces, can also provide insights into the broader functional architecture of visual processing. Traditional face processing models have often cantered on the higher-level, face-selective regions, sometimes overlooking the significance of V1/V2, even though these areas are foundational for visual processing. The bidirectional connections between low and high levels in the visual hierarchy, imply a potential role of V1 in sophisticated visual processes beyond low-level feature extraction. The investigation of these processes using complex stimuli can shed light on how visual information is integrated across different levels of the visual system.

In short, the investigation of recurrent processing - and the role of coarse-to-fine processing - in the visual system using complex stimuli and backward masking offers a novel approach to understanding the underlying mechanisms of visual perception. By examining the role of recurrent processing in the context of face photographs, I aimed to provide insights into the dynamic interplay between bottom-up and top-down processes, ultimately contributing to our understanding of visual cognition and its neural substrates.

Outline of the thesis

This thesis focuses on an in-depth examination of the role of recurrent processing when dealing with complex stimuli such as faces. Guided by two central lines of inquiry, the research first investigates the hypothesis that recurrent processing unfolds via a coarse-to-fine trajectory (addressed in questions 1 and 2). Subsequently, it probes the theory that recurrent processing intensifies in response to the degradation of stimuli (addressed in question 3). The investigation unfolds through the three following questions:

- "What is the role of the primary visual cortex (V1) in recurrent coarse-to-fine processing?"

Given that V1 has been theorised to function as an 'active blackboard' for integrating visual information in a coarse-to-fine manner (Bullier, 2001), it presumably holds a critical role in this integration process for complex visual stimuli. However, research is lacking in the exploration of V1's precise role in this process. To address this gap, chapter 2 describes an fMRI study, where we manipulate the visibility of different SF ranges through backward masking. This approach enables a critical examination of V1's engagement in refining visual information.

- "What are the behavioural implications of disrupting recurrent coarse-to-fine processing?"

Following the detection of V1 response modulations due to disrupted SF ranges and varied masking SOAs to be in line with coarse-to-fine models of processing, we became interested in understanding how these alterations translate to behavioural outcomes. Chapter 3 takes this inquiry further into the field of psychophysics, employing backward masking across diverse SF ranges. The aim is to address the impact of recurrent coarse-to-fine processing disruptions on perceptual behaviour.

- "How does stimulus degradation, specifically through a reduced signal-to-noise ratio, influence recurrent processing?"

Chapter 4 shifts the focus from SF manipulation to a closer look at stimulus degradation. Recognising that recurrent processing becomes more conspicuous when stimuli are ambiguous or degraded (e.g., O'Reilly et al., 2012; O'Reilly et al., 2013; Rajaei et al., 2019; Wyatte et al., 2012), this inquiry anticipates corresponding response changes in some regions identified in chapter 2's study. By combining ultra-high field fMRI with backward masking, we seek to understand how changes in stimulus visibility may affect recurrent processing in V1/V2.

The overarching aim of this thesis is to illuminate the role that recurrent processing plays into shaping our visual experiences. By utilising the technique of backward masking, particularly with complex stimuli such as faces, we strive to uncover new insights into the manner in which recurrent processing affects perception. This endeavour aligns with the broader goal of enhancing our comprehension of real-world visual perception mechanisms.

Methodology

In order to study the temporal dynamics of coarse-to-fine processing and the recurrent processing associated with degraded stimuli, we employed backward masking combined with fMRI for some of our studies. In this methodology section, I will concisely outline the utilisation of fMRI within our research framework, and elaborate on the statistical methodologies that were applied.

Functional Magnetic Resonance Imaging

Neuroimaging techniques have revolutionised the study of brain activity by providing non-invasive measurement methods. These techniques, such as fMRI, EEG, and MEG, operate on distinct principles, each offering its own advantages and limitations. While EEG and MEG excel in capturing rapid changes in brain activity with their high temporal resolution, their spatial resolution is relatively low. Controversially, fMRI exhibits exceptional spatial resolution but lacks temporal precision.

During an fMRI experiment, a subject is positioned within an MRI scanner and exposed to a strong static magnetic field generated by powerful magnets, such as 3 Tesla (3T) or 7T in our experiments. Radiofrequency pulses and gradients are applied to manipulate the alignment of hydrogen protons in water molecules, causing them to relax and emit radiofrequency signals. These signals are processed to generate images representing the spatial distribution of the MRI signal intensity. A core imaging technique in fMRI is the blood-oxygen-level dependent (BOLD) contrast, which leverages the magnetic properties of deoxyhaemoglobin to visualise areas of neural activity (Kwong et al., 1992; Ogawa et al., 1992, 1993). As neurons become active, the consequent rise in blood flow to the area not only meets but often exceeds the oxygen demand, reducing the concentration of deoxyhaemoglobin. This surplus of oxygenated blood relative to deoxygenated blood leads to an increase in the MRI signal, which the BOLD mechanism captures, offering an indirect indicator of neural activity. The hemodynamic response function (HRF) is central to understanding fMRI signals (Figure 1.7). It describes how blood flow changes in response to neural activity, illustrating the time course of the blood flow, volume, and oxygenation changes that occur in response to neural stimulation. The gradual nature of the hemodynamic response contributes to fMRI's lower temporal resolution, which poses a challenge in studying rapid cognitive events. Despite this limitation, fMRI's exceptional spatial resolution is invaluable for non-invasively mapping brain activity, particularly in understanding the spatial dynamics of cognitive processes.

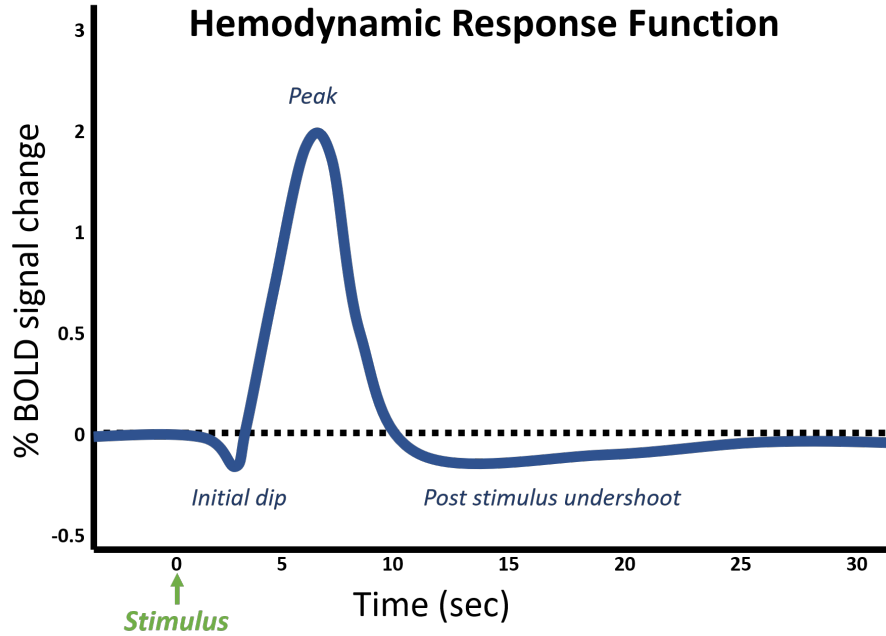


FIGURE 1.7: Model of the hemodynamic response function. The HRF is used in fMRI studies to represent the temporal changes in blood oxygenation and flow following neural activation. The graph shows the percentage change in BOLD signal over time in seconds following a stimulus, marked by the green line at the zero point on the time axis. The HRF begins with an initial dip representing a brief decrease in signal intensity. This is followed by a rapid increase as the blood flow to the activated brain region increases. After reaching the peak, the signal undergoes a gradual fall in intensity, reflecting the washout of excess oxygenated blood. Following this decline, there is a post-stimulus undershoot where the signal dips below the baseline before eventually returning to its initial level. This undershoot likely represents a complex interplay of factors, including the restoration of blood oxygenation and volume to their baseline states.

Backward masking can complement fMRI by providing the necessary temporal resolution to isolate fast occurring cognitive processes, such as recurrent processing. By manipulating the SOA, it is possible to exert precise control over the timing of perceptual events. This allows the dissection of early feedforward processing from later recurrent processing and can help establish a temporal order of activations across different brain regions. By leveraging fMRI, we aimed to precisely localise V1/V2 in each subject, enabling accurate examination of neural responses within these areas. To extract meaningful insights from the fMRI data, we employed specific analytical techniques, which will be elaborated upon in the subsequent sections.

General Linear Model

Both the fMRI data from studies described in Chapter 2 and Chapter 4 were approached in a similar manner following data pre-processing. A general linear model (GLM) was conducted to obtain quantified responses from different regions of interest. The GLM, a widely employed statistical framework for fMRI data analysis, allows for modelling and inference regarding the relationship between brain activity and experimental conditions or variables (Monti, 2011). Through this analysis, the objective is to assess whether, and to what degree, each predictor contributes to the fluctuations observed in the time-course of a voxel. Notably, such analyses have for example contributed to the identification of specialised regions involved in processing faces (e.g., Kanwisher et al., 1997), body parts (e.g., Downing et al., 2001) and scenes (e.g., Aguirre et al., 1998; Epstein and Kanwisher, 1998).

Upon collecting the fMRI data, pre-processing steps are performed to correct for various artefacts and noise sources. This typically involves procedures such as motion correction, spatial normalisation, and potentially spatial smoothing. Subsequently, the GLM necessitates the specification of a design matrix that describes the experimental conditions and their temporal characteristics. Each condition is modelled as a distinct predictor, represented by time-varying regressors indicating the onsets and durations of the conditions. To account for the delay between neural activity and the measured BOLD response in fMRI, the design matrix is convolved with a canonical HRF. This convolution generates the predicted BOLD response for each condition.

Parameter estimation is a critical step in the GLM, where model parameters are estimated by fitting the predicted BOLD response to the actual fMRI data. This estimation is commonly accomplished using least squares estimation or similar optimisation methods. The GLM utilises multiple linear regression to assess the impact of various predictors on the observed brain activity, following the equation:

$$Y = X\beta + \varepsilon$$

Here, Y is a column vector representing the time-series BOLD signal for a single voxel, X is the design matrix where each column denotes a different predictor variable, and β is the vector of unknown weights determining the unique association between each predictor variable and the data Y . The model also accounts for error values (ε), represented by a vector which contains the discrepancy associated with each observation, i.e., the value that remains unexplained by the weighted sum of predictor variables. Once the model parameters (beta values; β) are estimated, statistical tests can be performed to assess the significance of the effects of interest.

Psychophysiological interaction analysis

The univariate GLM analysis focuses on investigating hemodynamic response modulations within specific brain regions to infer their recruitment during task performance. While this method provides valuable insights into local information within individual regions, it does not reveal the flow of information between distant brain regions, i.e., their connectivity. One approach to analyse fMRI data and explore the relationship between brain activity in various regions and a psychological or behavioural variable is through a psychophysiological interaction (PPI) analysis. In PPI analysis, a specific brain region, known as the seed region, is selected. The goal of the analysis is to identify brain regions where the BOLD responses exhibit task-specific changes in functional connectivity with the seed region. The BOLD time-series from this region is multiplied by a psychological regressor, which represents the experimental task or condition of interest, to create the PPI regressor. Alongside this PPI regressor, the same psychological regressor and other nuisance variables (e.g., head motion) are entered into a multiple regression analysis, akin to the univariate GLM analysis. Regions showing a stronger connection with the seed region under one experimental condition compared to another are considered to be interacting with the seed region in a manner influenced by those conditions. This approach enables the study of how different tasks or conditions may influence the connectivity between brain regions, providing insights into the neural network dynamics underlying cognitive processes. PPI analysis is a powerful tool for examining the relationship between brain activity and experimental conditions. However, it is important to keep in mind that PPI analysis is a way to investigate connectivity, but does not inform on its directionality. Hence, it cannot be derived from this analysis whether the connectivity is due to feedback or feedforward signals. Nonetheless, PPI analysis offers valuable insights into which brain regions are coupled and play a role in recurrent processing of complex visual stimuli.

Linear Mixed effects Modelling analysis

In all experiments detailed in this thesis, the timing of the backward mask was manipulated with incremental SOAs. Additional experimental manipulations included varying the type of backward mask and stimulus type, with some studies also investigating different brain regions (Chapters 2 and 4). The data collected ranged from behavioural responses to fMRI beta values. A central focus across these studies was to investigate how response slopes were influenced by SOA. To analyse this data, I employed Linear Mixed effect Model (LMM) analysis. This statistical approach is especially advantageous for modelling relationships in data exhibiting correlations or dependencies, as is common in fMRI studies. Its flexibility and robustness make it suitable for our complex experimental design, allowing for the analysis of both fixed and random effects while accounting for the inherent correlations within the data.

In an LMM, the dependent variable is modelled as having a linear relationship with both fixed effects and random effects. Fixed effects represent factors or variables assumed to have a systematic and constant influence on the outcome variable. These effects are typically the main focus of the analysis and can be modelled as categorical or continuous predictors. For instance, in the experiment described in Chapter 2, we examined the impact of different backward masking conditions (e.g., LSF or HSF masks) on the brain response of each subject. These conditions were treated as fixed effects, with SOA being a continuous predictor and the other variables as categorical. On the other hand, random effects capture unexplained variability and account for correlations or dependencies within the data. These effects are often modelled as random intercepts or slopes, accommodating individual variations across different levels of a grouping variable, such as subjects in our experiments.

The LMM analysis concurrently estimates both fixed and random effects, while considering the covariance structure of the data. It enables the estimation of effects at both individual and group levels and quantifies uncertainty through standard errors and confidence intervals. LMM is also adept at handling unbalanced or missing data, making them suitable for complex study designs. The estimation of model parameters in LMM is typically performed using methods such as maximum likelihood estimation or restricted maximum likelihood estimation, and model fit can be assessed using metrics like the Bayesian information criterion (BIC). Overall, the LMM analysis has proven to be an essential approach for the studies in this thesis, offering nuanced insights into the complex interactions between different conditions (i.e., SF masks or SNR stimuli), and effectively capturing the underlying variability and dependencies in the recurrent processing of complex visual stimuli.

Chapter 2

Backward masking reveals coarse-to-fine dynamics in human V1

Corresponding publication:

*Schuurmans, J. P., Bennett, M. A., Petras, K., & Goffaux, V. (2023). Backward masking reveals coarse-to-fine dynamics in human V1. *NeuroImage*, 274, 120139.*

Jolien P. Schuurmans¹, Matthew A. Bennett², Kirsten Petras³, Valérie Goffaux^{1,2,4}

¹Psychological Sciences Research Institute, Université Catholique de Louvain, Louvain-la-Neuve, Belgium; ²Institute of Neuroscience, Université Catholique de Louvain, Louvain-la-Neuve, Belgium; ³Integrative Neuroscience and Cognition Center, CNRS, Université Paris Cité, Paris, France; ⁴Maastricht University, Maastricht, the Netherlands.

Abstract

Natural images exhibit luminance variations aligned across a broad spectrum of spatial frequencies (SFs). It has been proposed that, at early stages of processing, the coarse signals carried by the LSF of the visual input are sent rapidly from V1 to ventral, dorsal and frontal regions to form a coarse representation of the input, which is later sent back to V1 to guide the processing of fine-grained HSF. We used fMRI to investigate the role of human V1 in the coarse-to-fine integration of visual input. We disrupted the processing of the coarse and fine content of full-spectrum human face stimuli via backward masking of selective SF ranges (low SFs [LSFs]: $<1.75\text{cpd}$ and high SFs [HSFs]: $>1.75\text{cpd}$) at specific times (50, 83, 100 or 150ms). In line with coarse-to-fine proposals, we found that (1) the selective masking of stimulus LSF disrupted V1 activity in the earliest time window, and progressively decreased in influence, while (2) an opposite trend was observed for the masking of stimulus's HSF. This pattern of activity was found in V1, as well as in ventral (i.e. fusiform face area, FFA), dorsal and orbitofrontal regions. We additionally presented subjects with contrast negated stimuli. While contrast negation significantly reduced response amplitudes in the FFA, as well as coupling between FFA and V1, coarse-to-fine dynamics were not affected by this manipulation. The fact that V1 response dynamics to strictly identical stimulus sets differed depending on the masked scale adds to growing evidence that V1 role goes beyond the early and quasi-passive transmission of visual information to the rest of the brain. It instead indicates that V1 may

yield a 'spatially registered common forum' or 'blackboard' that integrates top-down inferences with incoming visual signals through its recurrent interaction with high-level regions located in the inferotemporal, dorsal and frontal regions.

Introduction

Visual perception results from the recurrent (i.e., iterative) communication within and across different levels of the cortical hierarchy via horizontal, feedforward, and feedback connections (Gilbert & Li, 2013; Hupé et al., 1998; Kravitz et al., 2013; Kreiman & Serre, 2020; Lamme et al., 1998; Petro et al., 2014). How recurrent processing converts the incoming visual input into a stable and rich perceptual experience, is still not completely understood (Olshausen & Field, 2006; Roland, 2010). One line of inquiry stems from the notion that visual function, therefore recurrent processing, has adapted to regularities in the spatial structure of the visual environment (over onto- and phylogeny; H. B. Barlow, 1975). One fundamental regularity is that natural images contain luminance variations over a wide spectrum of spatial scales, or frequencies (SFs) and that, across SFs, such luminance variations align in phase. Several authors proposed that recurrent processing exploits SF alignment to support the efficient integration of visual information (Bar, 2004; Bullier, 2001; Hegdé, 2008; Kreiman & Serre, 2020; Watt, 1987). In this framework, visual processing starts with the fast feedforward transmission of the coarse structure of the input conveyed by low SFs (LSF), presumably via the rapid and transient magnocellular pathway. Within the first tens of milliseconds of processing, this initial sweep activates a coarse representation of the visual input in high-level dorsal, frontal and ventral visual regions that retain prior knowledge about the natural statistics of the category best matching input (Bar, 2004; T. S. Lee, 2015; Mumford, 1992). These high-level coarse representations are then fed back to V1 (and possibly V2) and used to guide the processing of the fine-grained cues that are contained in high SF (HSF) input relayed via the slower and more sustained parvocellular pathway (Baseler & Sutter, 1997; Nowak et al., 1995; Schmolesky et al., 1998). V1 (and V2) is proposed to operate as the 'active blackboard' of the coarse-to-fine integration of visual input (Figure 2.1a). There are several hints that V1 yields such a 'spatially registered common forum' or 'blackboard' (Bullier, 2001; Deco & Lee, 2004; Roland, 2010) able to receive and integrate higher-order inferences with incoming visual signals via recurrent interactions from the rest of the brain. First, V1 encodes visual input with the greatest spatial specificity; it is thus necessarily involved for the sharp representation of the visual input (Bullier, 2001; Deco & Lee, 2004; T. S. Lee et al., 1998). Moreover, V1 neurons attune to progressively higher SFs over the course of temporal processing in line with a core tenet of coarse-to-fine proposals, namely that LSFs are encoded before HSFs (Allen & Freeman, 2006; Bredfeldt & Ringach, 2002; Einevoll et al., 2011; Frazor et al., 2004; Malone et al., 2007; Mazer et al., 2002; Nirody, 2014; Parker & Salzen, 1977; Tanaka & Sawada, 2022). Lastly, the fact that V1 is highly interconnected with the rest of the brain, with thalamic afferences thought to account for a very

limited fraction of V1 activity (Budd, 1998; Douglas & Martin, 2007; Kravitz et al., 2013; Markov et al., 2014; Muckli & Petro, 2013), indicates that responses in this region are largely modulated by non-retinal signals. Despite the increasing influence of coarse-to-fine theories in vision and computer sciences (for a review see e.g. Kreiman and Serre, 2020), and the fact that V1 is the best documented region in the primate brain, evidence for its contribution to human vision as an ‘active blackboard’ or ‘high-resolution buffer’ for coarse-to-fine integration is surprisingly scarce. Past research supports the notion that coarse input scales are processed before fine details (for simple stimuli such as lines, dots and gratings see e.g. Breitmeyer, 1975; Hughes et al., 1996; Jones and Keck, 1978; Mihaylova et al., 1999; Musselwhite and Jeffreys, 1985; Parker, 1980; Parker and Dutch, 1987; Watt, 1987 and for complex stimuli such as faces and natural scenes see e.g. Goffaux et al., 2011; Halit et al., 2006; Kauffmann, Chauvin, Guyader, and Peyrin, 2015; Kauffmann, Ramanoël, et al., 2015; Lu et al., 2018; Musel et al., 2014; Parker et al., 1992, 1997; Peyrin et al., 2006; Schyns and Oliva, 1994; Vlamings et al., 2009). However, because this research investigates LSF and HSF processing separately, via the use of SF-filtered stimuli, it cannot address a fundamental prediction of recurrent processing, i.e. that coarse-to-fine integration helps the visual processing of natural, full-spectrum, input. We recently addressed this question and showed that the availability of meaningful LSF information, as found in broadband phase-aligned images, indeed reduces the resources allocated to HSF processing in the (extra)striate cortex, compared to phase-scrambled images (Petras et al., 2019, 2021). This suggests that, in natural (full-spectrum) visual input where SFs align in phase and thus contain partially redundant information, the rapidly processed LSF information guides the processing of the rich and complex HSF, therefore reducing computational costs. Yet, the spatial resolution afforded by MEG (Petras et al., 2021; or EEG as in Petras et al., 2019) was insufficient to discern whether V1 itself is involved in the coarse-to-fine integration.

Here we systematically addressed the contribution of V1 to the coarse-to-fine integration of visual information. We reasoned that if coarse-to-fine integration is implemented in V1 for the sake of efficient processing, its disruption should induce a measurable cost in V1 processing (i.e., less sharp/sparse representations), which, in this region, typically results in an overall enhanced level of Blood-Oxygen level-dependent, i.e., BOLD, response (e.g., de-Wit et al., 2012; Freud et al., 2017; Goffaux et al., 2016; Kok and de Lange, 2014; Markov et al., 2014; Olshausen and Field, 1996; but see Rieger et al., 2013). We disrupted the coarse-to-fine recurrent processing of human face stimuli via backward masking, a method shown to block recurrent processing while preserving feedforward transmission (Bacon-Macé et al., 2005; Breitmeyer, 1984; Fahrenfort et al., 2007; Grill-Spector et al., 2000; Kovács et al., 1995; Lamme et al., 2002; Rolls & Tovee, 1994; Rolls et al., 1999). We showed full-spectrum face images for either 50, 83, 100 or 150ms; these stimulus onset asynchronies (SOAs) were chosen to track recurrent processing from its early stage on (see Ahmed et al., 2008; Eriksson and Roland, 2006; M. A. Harvey et al., 2009; Lamme, 1995; Roland, 2010; Roland

et al., 2006; Scholte et al., 2008; Takagaki et al., 2008; Wibrals et al., 2009). The mask contained either LSF or HSF noise, to selectively disrupt the coarse or fine processing of the face stimulus (Bacon-Macé et al., 2005; Gao & Bentin, 2011; Stromeyer & Julesz, 1972). This approach allowed us to disentangle the respective SF contribution to neural processing over time while accommodating a full range of aligned SFs in the stimuli (a critical aspect of naturalistic vision for the coarse-to-fine framework). We compared the backward masked responses between phase-aligned and phase-scrambled face stimuli in order to extract the visual responses specific to the processing of face shapes (Fahrenfort et al., 2007; Goffaux et al., 2011; see e.g. Grill-Spector et al., 2000 for a similar approach). The phase scrambling procedure maintained the amplitude spectrum of the original image while disrupting stimulus structure, resulting in unrecognisable images. Since intact and scrambled face stimuli were masked in a strictly identical way, our approach subtracted the direct influence of the mask on the BOLD response (Fahrenfort et al., 2007) and on the low-level mechanisms tuned to stimulus orientation/SF content. In unconstrained viewing conditions, V1 activity increases in response to stimulus scrambling (Figure 2.1b; Freud et al., 2017; Goffaux et al., 2016; Lerner et al., 2001), resulting in an overall negative intact-scrambled difference presumably due to an increase in the resources allocated to HSF processing (Petras et al., 2019, 2021). Following coarse-to-fine theoretical proposals, interrupting LSF processing early prevents top-down LSF-driven signals from guiding the formation of a sparse representation, therefore increasing V1 responses to the face stimulus and resulting in an almost null intact-scrambled difference. In line with the presumably decreasing contribution of coarse scales over time, LSF masking should cause progressively less disturbance over time. This is expected to result in a progressive decrease of the intact-scrambled differential response, until negative values are reached, indexing the successful formation of a face representation in V1. We anticipate the reverse tendency with the masking of HSF; during early processing stages, HSFs exert limited influence, resulting in minimal interference from their masking. As processing advances, the contribution of HSF processing to perception is hypothesised to increase, with disruptions from later masking amplifying the mismatch between feedback and feedforward information, thereby hampering efficient processing. This cross-like pattern of SF masking effects is the expected 'signature' of coarse-to-fine processing in V1 (Figure 2.1c). Similarly, high-level regions such as the fusiform, frontal and parietal areas, are thought to iteratively interact with V1 during coarse-to-fine integration, were also expected to exhibit this cross-like pattern; showing early LSF masking to be most disruptive and the opposite trend for HSF masking. Since these regions increase their response as a function of the input visibility (Kay & Yeatman, 2016; Yue et al., 2013), we expected an overall increase of their response as a function of SOA, irrespective of the masked SF. Still, we expected such increase to be steeper for HSF masking. The present design imposes a stringent test of the coarse-to-fine predictions as the stimuli are full-spectrum and identical across masking conditions so that any difference reflects the differential contribution of SF to vision over processing course. The coarse representation initially formed in high-level

visual regions has been proposed to involve some prior knowledge about the natural statistics of the category best matching input (Bar, 2004; T. S. Lee, 2015; Mumford, 1992). We thus expected contrast negation to hamper coarse-to-fine processing in V1 as this simple and reversible image manipulation disrupts the natural statistics of the human face stimulus (George et al., 1999; Liu-Shuang et al., 2022; Yue et al., 2013).

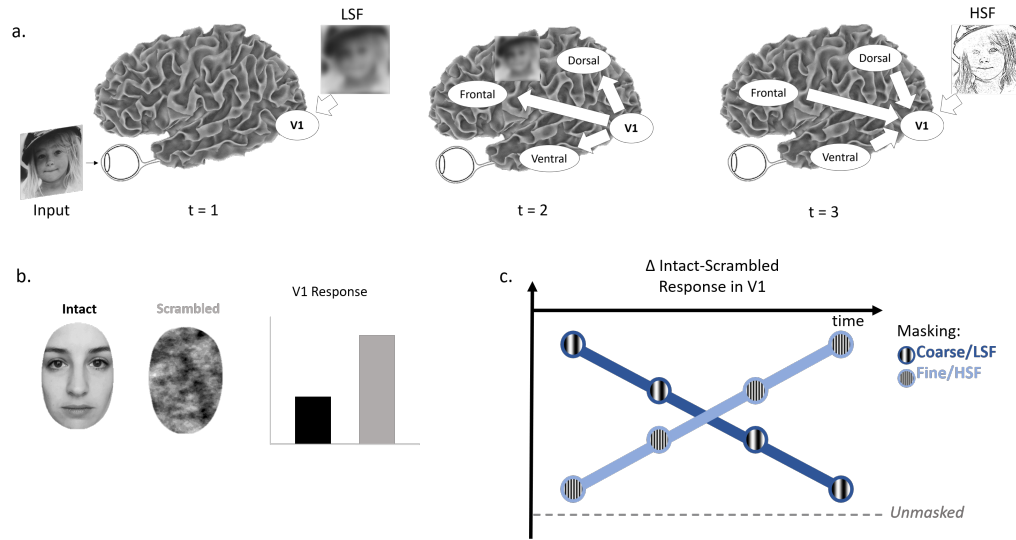


FIGURE 2.1: Expected responses of V1 to backward masking of selective SF ranges. a) The initial stage of visual processing involves rapid feedforward transmission of the coarse LSF input ($t=1$). This leads to activation of a coarse representation of the visual input in high-level dorsal, frontal, and ventral visual regions ($t=2$), which then feeds back to V1 for the subsequent processing of fine-grained HSF input ($t=3$). V1 is proposed to function as an "active blackboard" for the coarse-to-fine integration of visual input. b) In unconstrained viewing conditions, V1 response is lower for intact than scrambled stimuli. Contrasting these conditions results in an overall negative intact-scrambled difference; depicted as a grey dotted line in panel c. c) Consistent with coarse-to-fine processing theories, early masking of LSFs disrupts top-down LSF-driven signals, resulting in an almost null intact-scrambled difference. The disruptive effect of LSF masking is expected to decrease over time. The reverse tendency is expected for HSF masking, as the contribution of HSF processing to perception is hypothesised to increase over processing time. This cross-like pattern of SF masking effects is proposed as the "signature" of coarse-to-fine processing in V1.

Methods

Subjects

Eighteen healthy subjects (age 25.1 ± 3.28 , 5 males) - with no history of neurological disease - completed three scanning sessions each. Two subjects were excluded from data analysis due to inattentiveness and motion in the scanner. Before scanning, we tested the visual acuity of the subjects with the Freiburg Visual Acuity and Contrast Test (FrACT), which revealed all subjects had normal or corrected-to-normal vision (LogMar mean = -0.01 ± 0.11 , decVA = 1.08 ± 0.39). All subjects reported being right-handed as measured with the Edinburgh handedness Questionnaire (82.9 ± 17.3 ; Oldfield, 1971). Subjects were compensated financially, and a signed informed consent was obtained. The study was approved by the Ethical Committee of the Université Catholique de Louvain (CEHF 2016/13SEP/393).

Stimuli

Each subject performed the main experiment as well as a functional localiser to delineate the face-specialised voxels in the ventral pathway at the individual level. For both the functional localiser and main experiment, the stimuli were presented with PsychoPy v3.2.4 on a NordicNeuroLab LCD Monitor (32-inch, 1920 x 1080 pixels, 698.40 x 392.85 mm, frame rate = 60 Hz) situated at the end of the scanner bore. Subjects viewed the stimuli via a mirror attached to the head-coil, at a viewing distance of 175cm. Across experiments, all images were attributed a mean (i.e., the global luminance) and root mean square (RMS) contrast of 0.45 ± 0.1 .

Main experiment

We generated the main experimental stimuli based on 20 full-front greyscale photographs of unfamiliar human white faces (half male) with a neutral expression. Models were young adult alumni students (aged 18-25 years) of the Université Catholique de Louvain (Belgium), who gave written consent for the usage of their image (Laguerre et al., 2012). All faces were 367 pixels in height and 267 pixels in width and subtended a visual angle of $4.37^\circ \times 3.18^\circ$. To avoid the potential clipping of extreme intensity values in the image, we rescaled its luminance so that pixel values were between 0.1 and 0.9. Image luminance was then normalised (mean of zero and RMS contrast of 1). Next, faces were centred on a uniform grey background of 550 x 550 pixels ($6.54^\circ \times 6.54^\circ$ of visual angle).

Each face appeared in three variations: natural, contrast negated, and phase-scrambled. Negated faces were created by subtracting the face pixel values from 1. We displayed the faces in a phase-scrambled version as a control condition. Phase scrambling preserves the amplitude spectrum of the original image but disrupts shape and edge information and hence image recognisability. We scrambled images by randomising their phase structure in the Fourier domain using a custom MATLAB function. Since manipulations in the Fourier domain operate

on the whole image (face and background), the presence of a uniform background artificially increases the LSF energy of scrambled images. In order to minimise contamination of the phase-scrambled image by its originally uniform background, we employed an iterative phase scrambling procedure (as in Jacobs et al., 2020; Petras et al., 2019, 2021; see Figure 2.2a). Iterative scrambling involves first generating a phase scrambled version of the face image, then superimposing the original face back onto the scrambled image, scrambling this new image again, and repeating the procedure several times (for all our scrambled stimuli, 500 iterations were used). The iterative scrambling procedure results in an image in which the background and face share similar spectral properties. We employed iterative phase scrambling to create scrambled copies of the face stimuli ($n=20$), masks ($n=40$), and stimulus backgrounds ($n=10$).

We multiplied all stimuli and masks with a face-shaped aperture. Borders of this aperture gradually transitioned from 1 to 0 to avoid edge artefacts. A complementary aperture was multiplied with the scrambled background (see below) to avoid an increase of contrast in the transition region. To avoid abrupt stimulus on- and offset and instead provide a smooth transition between trial events within a block, all stimuli in a block were alpha-blended with a stable phase-scrambled background (e.g. see Ales et al., 2012; Figure 2.2b). Alpha blending creates a weighted sum of the stimulus and scrambled background contrast (a weight of 0.45 for stimulus and 0.55 for background) while maintaining mean luminance. The moderate RMS contrast of the stimuli and their reduced weight in the final alpha-blended stimulus were chosen to create challenging viewing conditions, and thus stimulate recurrent processing (Hupé et al., 1998; Kar & DiCarlo, 2021; Kar et al., 2019; Mohsenzadeh et al., 2018; Wyatte et al., 2014). As a result, the generated stimuli consisted of a face embedded in the phase-scrambled background (see Figure 2.2c). There were 20 blocks per condition, with the 10 phase-scrambled backgrounds repeated twice.

In contrast to the stimuli, the masks were superimposed on the background without alpha blending to ensure strong masking of the SF range it contains. Each stimulus (natural, negated, or scrambled face) was masked by a phase scrambled mask (Figure 2.2d). The scrambled masks were filtered to selectively contain either LSF or HSF, in order to selectively interfere with the processing of the LSF or HSF content of the full-spectrum face, respectively. In order to restrict mask image content to either LSFs or HSFs, we used 2nd order Butterworth filters with cut-offs such that four octaves were preserved for both SFs, without any gap between the two SF cut-offs (LSF: 0.7168 - 11.47cpi, HSF: 11.47 - 183.5cpi).

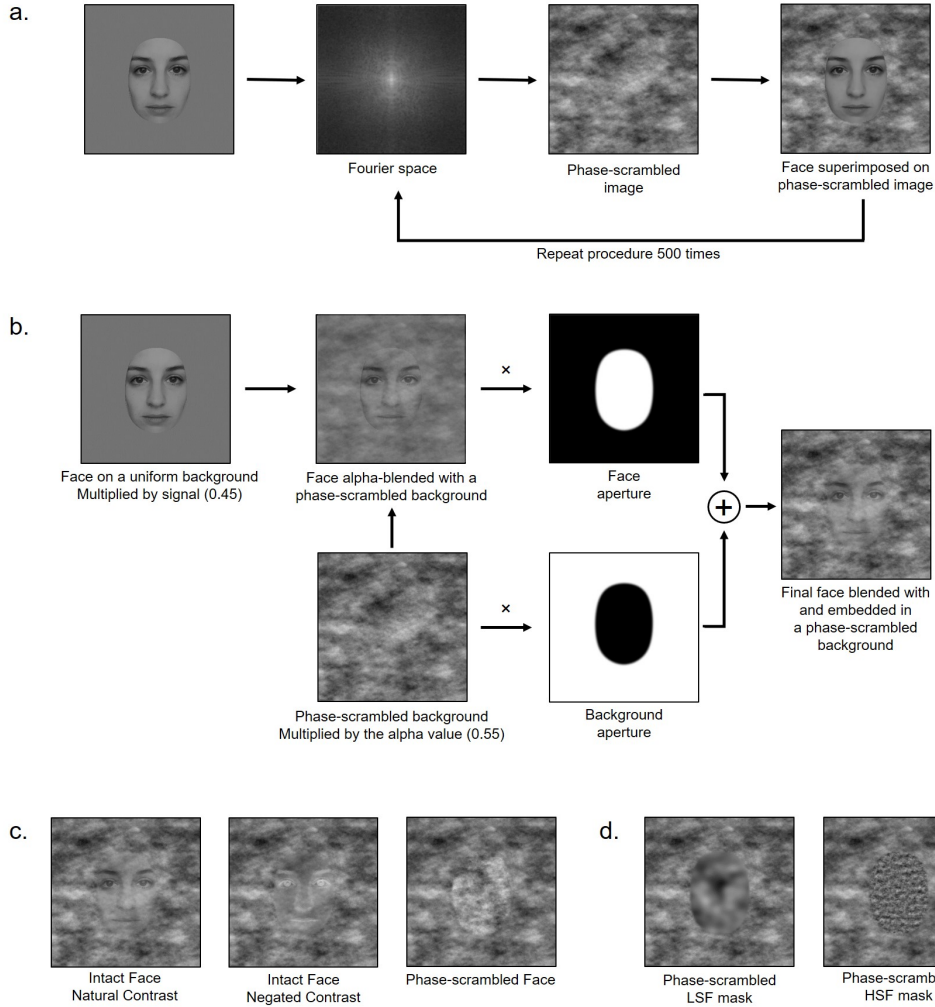


FIGURE 2.2: Stimulus generation. *a)* Iterative scrambling procedure. The image was first phase-scrambled in the Fourier domain, after which it was pasted back on the phase-scrambled image. This procedure was repeated 500 times until the final face-scrambled image had similar spectral properties as the face. *b)* To blend the face with the phase-scrambled background, we used alpha blending - an image combining method that relies on pixel transparency. Both the face image and background were multiplied with a face shaped aperture with a gradually transitioning border. The final face was embedded in and blended with the face scrambled background. *c)* All image types of an example face: intact natural and contrast negated stimuli, a phase-scrambled full-spectrum stimulus. The alpha blending was adapted for illustration purposes; experiment stimuli were less visible. *d)* Phase-scrambled masks containing either LSFs or HSFs. The masks were not alpha blended.

Functional localiser

The stimuli used for the localiser experiment were adapted from prior studies, available through the fLoc functional localiser package (Stigliani et al., 2015). Each of the three image categories - faces, hands, and instruments - contained 20 greyscale images as well as their phase-scrambled counterparts (see Supplementary material 1). The stimuli consisted of unfamiliar faces (half male) of various viewpoints (including hair), isolated hands in various poses, and stringed instruments (e.g. guitar, cello, lute, etc) positioned in different orientations. They were superimposed onto Fourier phase-scrambled backgrounds using the iterative scrambling procedure described above.

Procedure

All subjects participated in 20 experimental runs divided over three scanning sessions. The first session included a T1-weighted anatomical scan and the second session a functional face localising run. Experimental and localiser functional runs all had a block design, of 10-second blocks, alternated with 10 seconds of fixation with a 12-second fixation period at the beginning and end of the run. A fixation period consisted of a uniform grey screen. Throughout all runs, a fixation cross was visible, consisting of two thin black lines connecting opposite corners of the square stimuli. Subjects were instructed to fixate on its central intersection at all times.

To ensure subjects paid attention during functional runs, we instructed them to detect a rare and brief colour change of the stimulus by pressing a button with the right index finger. In each block, there were two targets (red for the main experimental runs: HSV profile = 0.8, 1.0, 1.0, blue for the functional localiser run: HSV profile 1.0, 1.0, 0.8). One colour change occurred per block half, but never during the first stimulus of a block.

Main experiment

In each 8.7-minute run (see Figure 2.3a), there were 24 experimental conditions (3 stimulus types by 2 SF masks by 4 SOAs) presented in separate blocks and two retinotopic localiser conditions. The order of experimental blocks was randomised with the constraint that no condition followed any other particular condition more than three times during the course of the experiment. The retinotopic localiser conditions at the end of every run consisted of a single block of 4Hz contrast-reversing checkerboard stimuli: in one condition limited to the region where the face appeared in the experimental conditions and the other to the background. This was used to localise the region of the (extra)striate cortex responding to the stimulus region containing the face selectively. Each block included 20 stimulus trials of the same condition (i.e., a given stimulus type in a given SOA and SF mask condition) and four pseudo trials (417ms each; 25 frames at 16.67ms per frame) in which the background remained but no face or mask was presented. Two pseudo trials appeared at random during the first half of a block, and two

more appeared at random during the second half. These were included to make the appearance of the stimuli less predictable. The other 20 trials contained stimuli from the same experimental condition (i.e. one stimulus type, SF mask, and SOA per block). The order of the 20 stimulus trials in a block was randomised with the constraint that each trial was equally likely to appear at the beginning, middle and end of the block. The background remained stable within the block. A stimulus trial consisted of a face stimulus presented for either 50, 83, 100 or 150ms, followed by a 166ms mask, after which only background remained (see Figure 2.3c). Each event duration within a trial was a multiple of the frame rate (16.67ms).

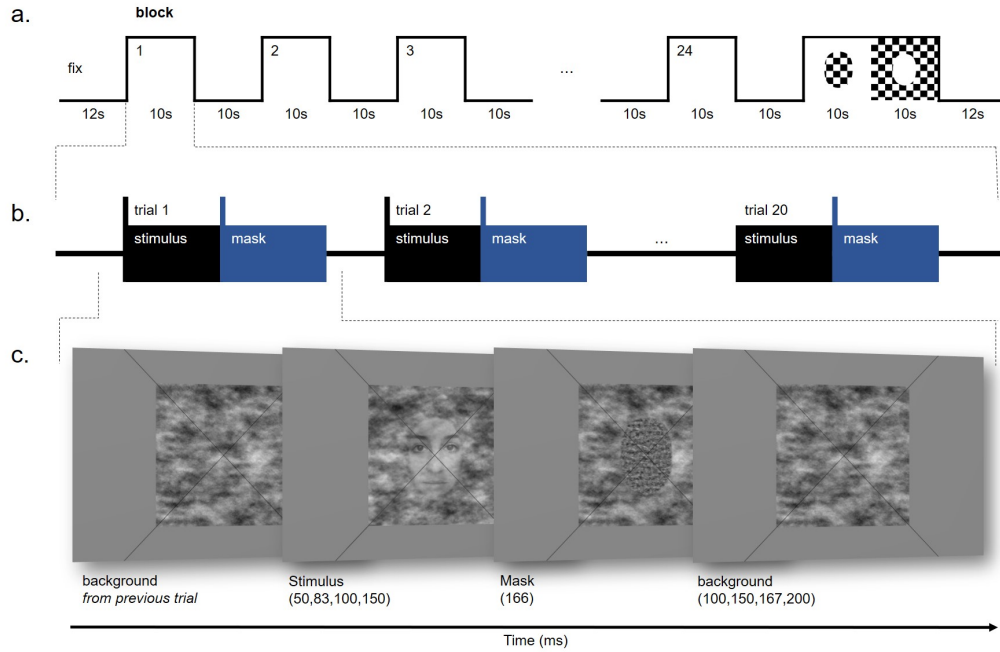


FIGURE 2.3: Design of main experiment. a) Every experimental run contained 24 blocks, alternated with fixation periods and ended with two retinotopic checkerboard blocks. b) A block consisted of 20 stimulus trials of the same experimental condition (i.e. a single stimulus type, SF mask, and SOA). c) An example trial: a face stimulus followed by a phase scrambled mask only containing HSFs. The alpha blending of the face stimulus was adapted for illustration purposes; experimental stimuli were less visible.

Functional localiser

During the 12.2-minute localiser run, subjects viewed greyscale images of either intact or scrambled faces, hands or instruments. Images from all categories were presented in 36 blocks in total (one image category per block), with six blocks per condition. A condition was never repeated twice in a row and the order of conditions was counterbalanced across subjects. A block comprised of 10 images in a random sequence, and each image was repeated three times throughout the functional localiser run. Each image appeared for 500ms, followed by a 500ms interstimulus interval.

MRI acquisition

Subjects were scanned in a 3Tesla GE Signa Premier MRI scanner with a 48ch head coil, based at Cliniques Universitaires UCL Saint-Luc in Brussels. As anatomical references, whole-brain T1-weighted images were obtained during the first sessions (3D MP-RAGE, 1 x 1 x 1 mm, FOV = 256 mm, TI = 900ms, FA = 8°). Functional T2*-weighted GE echo-planar imaging was used to obtain the BOLD signal as an indirect measurement of neural activity. Thirty-two 2.4-mm axial slices were acquired (2.4mm isotropic, FOV = 240mm, TR = 2000ms, TE = 30ms, FA = 90°).

Preprocessing

Functional and anatomical data were organised into the Brain Imaging Data Structure (Gorgolewski et al., 2011). Pre-processing of the data was carried out with fMRIPrep 20.1.1 (Esteban et al., 2018, 2019), which is based on Nipype 1.5.0 (Gorgolewski et al., 2011, 2016). To ensure reproducibility using the specific software versions for fMRIPrep and all its dependencies, it was executed via its Docker container (Merkel et al., 2014).

Anatomical data pre-processing

Each T1-weighted (T1w) volume was corrected for intensity non-uniformity (Tustison et al., 2010 *N4BiasFieldCorrection*), and skull-stripping was executed (*antsBrainExtraction.shv*, OASIS30ANTs template). Next, brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and grey-matter (GM) was performed on the brain-extracted T1w (using fast from FSL 5.0.9; Y. Zhang et al., 2001). Finally, brain surfaces were reconstructed (using recon-all from FreeSurfer 6.0.1; Dale et al., 1999), and the brain mask estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and *FreeSurfer*-derived segmentations of the cortical grey-matter of *Mindboggle* (Klein et al., 2017).

Functional data pre-processing

For each of the 24 functional runs per subject (across all tasks and sessions), the following pre-processing was performed. First, to generate a functional reference, volumes with substantial T1w contrast derived from non-steady states of the scanner (volumes at the beginning of EPI sequence) were identified, realigned, and averaged. After skull-stripping of the functional reference volume, head motion parameters with respect to the functional reference (transformation matrices, and six corresponding rotation and translation parameters) were estimated (Jenkinson et al., 2002; *mcfliirt* - FSL 5.0.9). On average, the maximum movement was $1.7 \pm 0.32\text{mm}$. After correcting for slice timing (Cox and Hyde, 1997; *3dTshift* from AFNI 20160207), the functional reference was co-registered to the T1w reference using boundary-based registration (*bbregister*, *FreeSurfer*; Greve and Fischl, 2009). The co-registration was configured with six degrees of freedom (i.e. 3 rotations and 3 translations).

Next, all functional data were resampled onto their original, native space by applying the transforms to correct for head-motion. Several confounding time series were calculated based on the functional data: framewise displacement (FD), DVARS and three region-wise global signals. FD was computed with the absolute sum of relative motions (*Power*, Power et al., 2014) and the relative root mean square displacement between affines (*Jenkinson*, Jenkinson et al., 2002). FD and DVARS were calculated for each functional run, both using their implementations in *Nipype* (Power et al., 2014). Frames that exceeded a threshold of 0.5 mm FD or 1.5 standardised DVARS were annotated as motion outliers. The three global signals were extracted within the CSF, the WM, and the whole-brain masks.

Using *FEAT* (FMRI Expert Analysis Tool; Version 6.0; FMRIB's Software Library, www.fmrib.ox.ac.uk/fsl) the functional data were smoothed in the space domain using a Gaussian kernel of FWHM 5mm. And high-pass temporal filtering was carried out using a Gaussian-weighted least-squares straight-line fitting ($\sigma = 50.0\text{s}$). A General Linear Model (GLM) was applied per voxel to capture each condition through a regressor, set with block onset and a duration of 10 seconds, and convolved with a hemodynamic response function. In total, 33+ regressors were generated, including 24 conditions (3 stimulus types by 2 SF masks by 4 SOAs), fixation, and 2 checkerboard conditions. To account for nuisance factors, 6 motion parameters and motion outliers (if present) were also included in the model as regressors. Contrasts were generated for all 16 intact conditions (8 in natural contrast and 8 negated contrast) by comparing them to the 8 scrambled conditions. These contrasts were corrected for family-wise error and considered statistically significant at $p < 0.5$. The GLM contrasted parameter estimates (i.e. beta-values) were obtained from the model.

For the whole-volume group-level analysis, we transformed for each subject the first-level estimate maps intact-scrambled for each condition into standard MNI152NLin2990c space with a resolution of 2mm3 using ANTs (*antsApplyTransforms*) employing the BSpline interpolation method. After the transformation, we smoothed the data with a gaussian kernel (FWHM = 8mm) using the *fslmaths* command.

Regions of interest

V1 was delineated based on anatomical landmarks and functional data. We first localised V1 anatomically based on the *Freesurfer* atlas label (Figure 2.4c). Next, we selected the voxels showing a significantly stronger response to the stimulus region containing the face compared to its background (following the procedure of de-Wit et al., 2012; F. W. Smith and Muckli, 2010). To do so, we contrasted the responses to the flickering checkerboard conditions in the block at the end of every run: [stimulus- background] (see Figure 2.4a). The significant voxels ($t > 1.65$, uncorrected $p < .05$) from this contrast (Figure 2.4b) that fell within V1, after smoothing, were defined as the V1 region of interest (ROI) in the current study (Figure 2.4d and Table 2.1). Follow-up analyses indicate that a more conservative thresholding approach yields similar results (see Supplementary material 2).

The region responding preferentially to faces in the fusiform cortex (FFA) was identified independently for each subject, based on the functional localiser scan. First, to select the brain regions more responsive to faces than non-face objects we computed the conjunction between [intact faces - intact hands] and [intact faces - intact instruments] contrasts (Figure 2.4e). Next, to exclude activity merely driven by the amplitude spectrum of face images we selected the voxels that showed a significantly larger response to intact than scrambled faces [intact face - scrambled face] (Rossion, Caldara, et al., 2003). Significant voxel clusters on t maps were selected at a q [false discovery rate, FDR] < 0.01 . After visual inspection, the threshold for some subjects was increased (cluster-threshold z -value > 3.0) to delineate between the fusiform and occipital face preferring clusters (see Figure 2.4f for an example subject). Since we were not able to delineate mid and posterior fusiform areas in all subjects, we did not distinguish between mFus- and pFus-faces in the current study (e.g. Weiner and Grill-Spector, 2013). Overall, and in line with past evidence (Kanwisher et al., 1997; Rossion et al., 2000), there were more face-selective voxels in the right FFA compared to the left (see Table 2.1). The voxel coordinates were in correspondence with previous research (see Table 2.1; Fahrenfort et al., 2012; Goffaux et al., 2011; Kanwisher et al., 1997; Rossion, Schiltz, and Crommelinck, 2003; Thome et al., 2021).

Additionally, we used the data obtained from the functional localiser to confirm that voxels in V1 showed greater responses to scrambled compared to intact faces (Goffaux et al., 2016). To test whether the [intact face - scrambled face] contrasted beta-values in V1 were significantly differing from 0, we conducted one-sample t -tests with a Bonferroni multiple comparison correction.

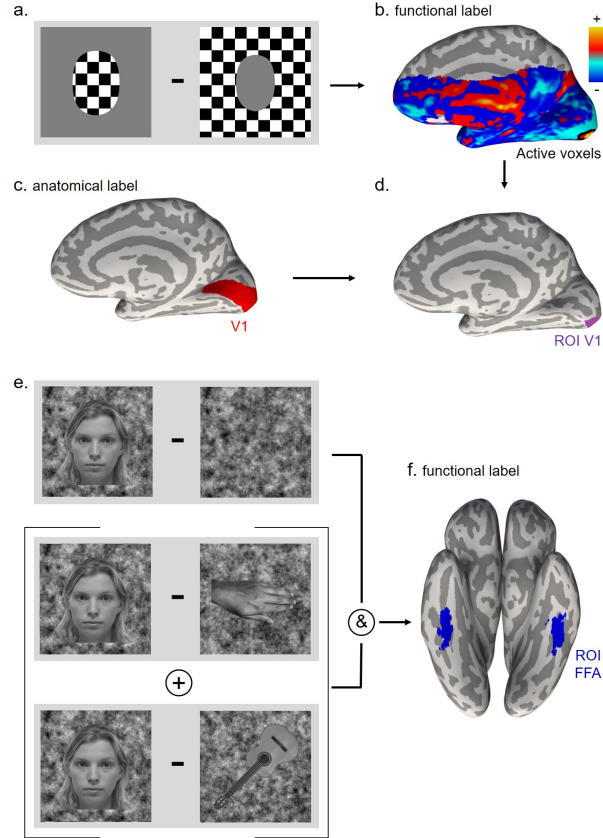


FIGURE 2.4: V1 and FFA ROI selection. a) Example checkerboard patterns of the retinotopy localiser. The pattern was confined to either the region where the face appeared in the main experiment or to the background. b) Contrasted [face checkerboards - background checkerboards] activation maps projected on an inflated brain surface of an example subject. A line is drawn to delineate the cluster of activated voxels in the (extra)striate cortex that show significantly stronger activation to face checkerboards than to background checkerboards ($p < .05$). c) The region of V1 was delineated using the Freesurfer atlas label. d) Inflated brain surface of an example subject with the final ROI consisting of voxels that fell both in the stimulated region (b) and the Freesurfer V1 region (c). e) Functional localisation stimuli examples. To rule out activity driven by the amplitude spectrum of face stimuli, we contrasted them with phase-scrambled stimuli. We then computed the conjunction of this cluster with a cluster more active to faces compared to hands or instruments. This way we ensured that we selected the brain region that is more receptive to faces than non-facial stimuli (Rossion et al., 2012). f) Example subject with the functionally defined FFA projected on the surface area of an inflated brain.

TABLE 2.1: ROI MNI152 coordinates and volume sizes averaged over 16 subjects (with standard deviation). Coordinates are based on the median voxel in each ROI.

ROI	hemi	N voxels	x	y	z
FFA	Right	143(104)	41(3.4)	-47(8.4)	-19(2.9)
	Left	108(93)	-27(2.8)	-52(9.4)	-14(3.1)
V1	Right	144(65)	14(5.2)	-91(7.1)	-3(5.1)
	Left	155(53)	-11(6.0)	-93(2.4)	-8(4.1)

Statistical Analyses

Behaviour analysis

In order to maintain attention to the stimulus throughout the experiment, subjects had to indicate whether the entire stimulus turned slightly red. Because of technical difficulties in the experimental code, the RTs were not correctly recorded for all trials and could not be analysed. We determined the accuracy for each condition per subject, ran a Linear Mixed effects Model (LMM) in R, and used a *Type III Analysis of Variance* (ANOVA; *Satterthwaite's* approach) to determine whether there were differences in accuracy between conditions. We used a step-wise variable selection procedure with backward elimination (Heinze et al., 2018) and ended up with a model only including main effects ($BIC = -1443.545$):

$$Accuracy \sim 1 + SOA + SFmask + StimulusType + (1|subject)$$

ROI analysis

GLM Individual beta weights were estimated for each condition of the SF experiment from each average ROI time course (left and right V1 and FFA). To separate the activity related to the processing of the face stimulus from general visual responses, we contrasted intact-scrambled in each experimental condition. These contrasted beta-values were used for the second-level analysis.

For each ROI in the second-level analysis (conducted in R), we used LMM approach to fit the contrasted beta-values from the first-level GLM, with fixed effects of SOA, SF masking, stimulus type, hemisphere and their interactions. Global variability between subjects was accounted for by modelling subject as a random effect (allowing individuals to have their own intercept about the group mean). While the LMM approach provides a robust framework for incorporating both fixed and random effects, it is important to acknowledge that the reliance on fixed effects for factors such as SOA, SF masking, and stimulus type, may not fully capture the random variability across individual subjects beyond their intercepts. This limitation could affect the generalisability of our findings to a broader population. For model fitting, we used the stepwise variable selection procedure, following backward elimination (Heinze et al., 2018). We started with the most complex model containing all variables and interaction terms (full model;

outcome of the full models were used to plot the data for illustrative purpose), then we removed the non-significant four, three, and two-way interactions one by one to find the simplest model possible and thus maximising power in the analysis (for a detailed step-by-step overview of the fitted models, see Supplementary material 4). We validated the resulting model by determining whether it yields the smallest Bayesian information criterion (BIC).

In both ROIs, there was no significant effect of hemisphere, so we collapsed the data from the two hemispheres. The final model for V1 included a two-way interaction term of SF masking and stimulus type, as well as SF masking and SOA (BIC = 118624.6):

$$BetaValue \sim 1 + SOA * SFmask + SFmask * StimulusType + (1|subject)$$

The final model for FFA included two-way interaction terms of SOA with stimulus type, and SF masking with SOA (BIC = 107475):

$$BetaValue \sim 1 + SOA * SFmask + SOA * StimulusType + (1|subject)$$

To interpret the estimated means we used the *"stats"* package in R to run a *Type III ANOVA* using *Satterthwaite's* method for degrees of freedom. Since we expected to see the cross-like pattern of differential activity, i.e. high differential responses for the early disruption of LSF, decreasing as a function of SOA and the opposite for HSF masking, we obtained slope values for the different conditions (with the *emtrends* function of the *emmeans* package). The difference in the slope values was then tested using a pairwise comparison approach (*pairs*). It is important to note that these final tests considered only the fixed effects and their interactions. Given the negligible contribution of random effects to our model, these tests did not explicitly incorporate aspects of subject variability.

Whole-volume group-level analysis

To identify other regions involved in coarse-to-fine processing, we performed a whole-volume group-level analysis, again using a LMM approach to fit the contrasted beta-values from the first-level GLM (i.e., outcome of intact-scrambled contrast). Here, we fit the same model on every voxel in the fMRI data and, unlike the models fit for the ROIs, retained all interaction terms regardless of whether or not they were significant. This is because different voxels would have required different terms to be removed and thus reduced the interpretability of the resulting statistical map (since each voxel would have been modelled differently). Therefore, we opted in favour of interpretability of the results (possibly sacrificing power for lower order interaction terms in models with redundant higher order interaction terms). For every voxel, we fit the full LMM (using *R*) including all interaction effects:

$$BetaValue \sim 1 + SOA * SFmask * StimulusType + (1|subject)$$

We performed an ANOVA on each voxel’s model predictions to obtain a p-value for each term, resulting in p-value maps. We adjusted the p-value maps for multiple comparisons using FDR correction (in *R*, using the *p.fdr* function from the *FDRestimation* package with `adjust.method="BH"`). To ensure that the voxels with an SOA by SF masking interaction effect indeed exhibit the ‘coarse-to-fine signature’ (i.e. more positive slopes when accumulating HSF than LSF), we obtained the slope values for the LSF and HSF masking conditions as a function of SOA. We subtracted the LSF from the HSF slopes, such that positive values support the ‘coarse-to-fine signature’. There were no voxels with a negative slope difference (i.e., no significant ‘fine-to-coarse signature’). To visualise the results on the cortical surface, we projected the difference map of the fitted slopes for LSF and HSF and thresholded this map with the FDR corrected p-value map.

Psychophysiological Interaction

To address whether the activity between V1 and FFA is coupled more strongly when coarse-to-fine processing can take place compared to when we disrupt it via masking, we conducted a psychophysiological interaction (PPI) analysis. The average time-series of the V1 ROI was used as the seed and was extracted using *fslmeants* to form a physiological regressor. A psychological regressor was created for each condition by convolving a double gamma HRF with the task regressor (O’Reilly et al., 2012). The PPI regressors were created by multiplying the physiological and psychological regressors (one per condition). The model also incorporated the six motion regressors. For all subjects, the different regressors were entered into a GLM, and the PPI interaction terms were estimated in the FFA.

Next, we applied the LMM stepwise variable selection procedure with backward elimination, similar to the main analysis. Starting with the most complex model, and removing non-significant interaction terms to yield the simplest model that accounted for the data ($BIC = 28482.8$):

$$BetaValue \sim 1 + StimulusType + (1|subject)$$

Results

Behavioural results

Subjects performed an easy colour change detection task, which randomly occurred twice per block. The performance accuracy was 97.02% (SD = 0.009). After fitting an LMM with full interaction effects and pruning this model to the simplest model with only main effects, the ANOVA showed that there were no significant differences of accuracy between any of the conditions (p-values ranging between .234 and .588; Table 2.2). The global variance between subjects only explained a small proportion of the residual variability (0.1%).

TABLE 2.2: Type III ANOVA - for accuracy of behavioural responses and ROI data in V1 and FFA.

	Sum Sq	Mean Sq	NumDF	DenDF	F	p
			Accuracy			
SOA	0.00000005	0.00000005	1	364	0.0000	.996
SFmask	0.00013665	0.00013665	1	364	0.0648	.799
StimulusType	0.00196569	0.00098284	2	364	0.4660	.628
			V1			
SOA	3110	3110	1	10187	0.4808	.489
SFmask	154801	154801	1	10187	23.9336	<.001
StimulusType	7911	7911	1	10187	1.2231	.269
SOA:Sfmask	192966	192966	1	10187	29.8342	<.001
SFmask:StimulusType	21943	21943	1	10187	3.3926	.066
			FFA			
SOA	1273009	1273009	1	10187	589.4631	<.001
SFmask	52022	52022	1	10187	24.0887	<.001
StimulusType	2	2	1	10187	0.0009	.976
SOA:Sfmask	58632	58632	1	10187	27.1492	<.001
SOA:StimulusType	66649	66649	1	10187	30.8614	<.001

ROI results

V1

First, to verify that the intact-scrambled response difference was negative in less constrained (i.e., unmasked and less rapid) stimulation conditions, we compared responses of intact and phase-scrambled faces from the functional localiser in V1. We found that the intact-scrambled response difference ($M = -.16$, $SD = .26$) was significantly lower than zero ($t(31) = -3.56$, $p = 0.002$).

In the main experiment, we predicted that the early masking of LSF would result in a quasi null intact-scrambled difference, in line with the hypothesis that early LSF processing guides the formation of a sharp representation in V1, and preventing such guidance results in a heightened response to intact faces. However, at 50 ms of SOA, the intact-scrambled difference was positive (see Figure 2.5) indicating that the response in V1 was larger for intact than for scrambled faces. Although this may appear puzzling, it could reflect the stimulus heterogeneity of intact blocks, and thus stronger stimulus on- and offset responses than in scrambled blocks. LSF masking caused progressively less disturbance over time as shown by the progressive decrease of the intact-scrambled differential response. At 150 ms, the response differential reached negative values, indexing the successful formation of a face representation in V1. The reverse pattern was observed when masking HSF, creating the hypothesised cross-like pattern of SF masking, the expected 'signature' of coarse-to-fine processing in V1. These observations were borne out by the LMM (Figure 2.5).

After systematically removing non-significant higher-order interaction terms (as determined by a *type III ANOVA*), the LMM included two-way interaction effect terms of SF masking with SOA and of SF masking with stimulus type (see Table 2.2). Only a negligible proportion of the residual variability was accounted for by the global differences between subjects (0.001%). Table 2.2 summarises the significant terms in the LMM. Since we did not find any hemispheric differences, we collapsed the data over both hemispheres in the LMM for more power in the analysis. We found a significant interaction effect between SOA and SF masking, and a marginally significant interaction effect between SF masking and stimulus type. Figure 2.5 depicts the average BOLD activity and the LMM predictions from the full model, although not all terms were significant we include them in the figure for illustrative purposes.

For both naturally contrasted and negated images, we found that in V1 the intact-scrambled response differential depicted a cross-like pattern. There was a negative slope as a function of SOA when masking LSFs (collapsed over stimulus type: $\beta = -.14$, $CI = .12$), whereas HSF masking yielded a positive slope ($\beta = .11$, $CI = .12$). These slopes differed significantly from each other ($z = -5.46$, $p < .001$).

Table 2.3 shows the estimated marginal means of the model as well as pairwise comparisons between LSF and HSF masked conditions. V1 response differential was significantly bigger when disrupting LSFs compared to HSFs at 50ms. By 150ms, this trend was reversed with disruption of HSFs causing a significantly larger V1 response than disruption of LSFs (see Table 2.3).

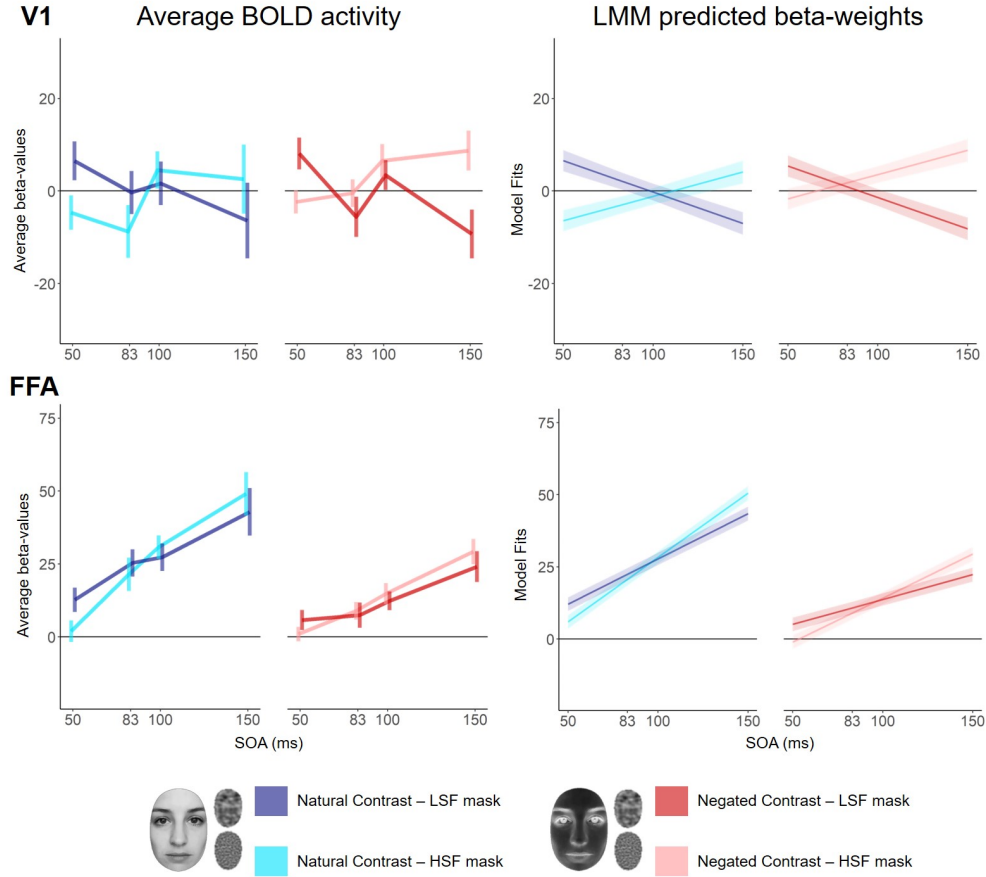


FIGURE 2.5: Intact-scrambled response differential for both V1 (top) and FFA (bottom). Plots in the left column represent the group averaged beta-values including a 95% within-subjects confidence interval as described in Loftus and Masson (1994). The LMM fits shown on the right plots are from a full model including all interaction effects which was ran just for illustrative purposes. In both V1 and FFA, there was a significant interaction effect of SOA with SF masking. The FFA showed an additional interaction effect of SOA with stimulus type.

FFA

As with V1, the proportion of residual variance explained by the random effect of subject was negligible (0.028%) and there was no interaction nor main effect of hemisphere; we therefore collapsed the data over left and right FFA. The LMM in FFA included the interaction effect between SOA and stimulus type, as well as between SOA and SF masking.

Unlike V1, the intact-scrambled response differential was positive in the FFA as this region is known to respond more to intact faces than scrambled faces (e.g. Rossion et al., 2012). Irrespective of the masked SF, the differential intact-scrambled response increased as a function of SOA. The influence of SOA was stronger for naturally contrasted faces ($\beta = .38$, $CI = .07$) than for contrast negated faces ($\beta = .24$, $CI = .07$). The slopes differed significantly ($z = -5.56$, $p < .001$), and there was a significant interaction between SOA and stimulus type (see Table 2.2 and Figure 2.5).

We observed the hypothesised coarse-to-fine signature in the form of a significant interaction between SOA and SF masking; similar to V1, the fitted slope was more positive when masking HSF ($\beta = .38$, $CI = .07$) than LSF ($\beta = .21$, $CI = .07$; test of the slope difference: $z = -5.21$, $p < .001$).

TABLE 2.3: Model predictions (estimated marginal means - emmeans) and pairwise comparisons on predicted variables between LSF and HSF masked conditions for different SOAs in both V1 and FFA.

ROI	SOA	LSF mask emmean (SE)	HSF mask emmean (SE)	Estimate (SE)	Test statistic (z/t)	p
V1						
			Natural			
	50	6.56 (2.28)	-6.40 (2.28)	13.00 (3.02)	4.29	<.001
	83	2.08 (1.82)	-2.93 (1.82)	5.02 (2.32)	2.16	.031
	100	-0.22 (1.79)	-1.15 (1.79)	0.92 (2.26)	0.41	.683
	150	-7.01 (2.46)	4.11 (2.46)	-11.10 (3.28)	-3.38	<.001
			Negated			
	50	5.39 (2.28)	-1.71 (2.28)	7.10 (3.02)	2.35	.019
	83	0.91 (1.82)	1.76 (1.82)	-0.85 (2.32)	-0.37	.715
	100	-1.40 (1.79)	3.55 (1.79)	-4.94 (2.26)	-2.19	.029
	150	-8.18 (2.46)	8.80 (2.46)	-17.00 (3.28)	-5.17	<.001
FFA						
			Natural			
	50	12.11 (2.36)	5.95 (2.36)	6.16 (1.48)	4.15	<.001
	83	22.46 (2.15)	20.68 (2.15)	1.78 (0.98)	1.82	.068
	100	27.79 (2.13)	28.27 (2.13)	-0.48 (0.93)	-0.517	.605
	150	43.47 (2.44)	50.59 (2.44)	-7.11 (1.66)	-4.29	<.001
			Negated			
	50	5.11 (2.36)	-1.05 (2.36)	6.16 (1.48)	4.15	<.001
	83	10.79 (2.15)	9.01 (2.15)	1.78 (0.98)	1.82	.068
	100	13.72 (2.13)	14.2 (2.13)	-0.48 (0.93)	-0.517	.605
	150	22.32 (2.44)	29.44 (2.44)	-7.11 (1.66)	-4.29	<.001

Whole-volume group-level results

In a whole-volume group-level LMM analysis, we investigated potential coarse-to-fine dynamics outside of V1 and FFA. The resulting p-value maps were FDR corrected for multiple comparisons. The only 2-way interaction term that showed significance in the whole-volume analysis was that of SOA and SF masking and there were no voxels with significant 3-way interactions. Several regions showed this interaction (see Figure 2.6 and Table 2.4). Along the ventral visual stream, the interaction was found significant in the right fusiform gyrus (BA 37), confirming findings from the ROI analysis. We did not find the effect to be significant in the left fusiform gyrus. The SOA by SF masking interaction was also significant in bilateral parietal and ventrolateral/orbitofrontal clusters (e.g., middle temporal gyrus or BA 21; inferior temporal gyrus or BA 20; see Table 2.4) - which have been theorised to be an origin of the coarse template-based feedback in coarse-to-fine processing (e.g., Bar, 2004; Bullier, 2001). In all cases, these 2-way interactions were due to a pattern reminiscent of what we already observed in the ROI analyses for V1 and FFA (Figure 2.5): a more positive slope with the progressive accumulation of HSF compared to LSF (HSF-masking slope: $\beta = .062$, $CI = .002$; LSF-masking slope: $\beta = -.089$, $CI = .002$; test of the slope difference: $z = -197.71$, $p < .001$). Specifically, at 50ms the intact-scrambled response differential was larger when disrupting the LSFs compared to HSFs, with the reverse being true at 150ms, in line with the anticipated coarse-to-fine signature. The effect of stimulus type (faces with natural vs negated contrast) did not survive the multiple comparisons correction ($F_s < 9.35$, $q[FDR] > .95$).

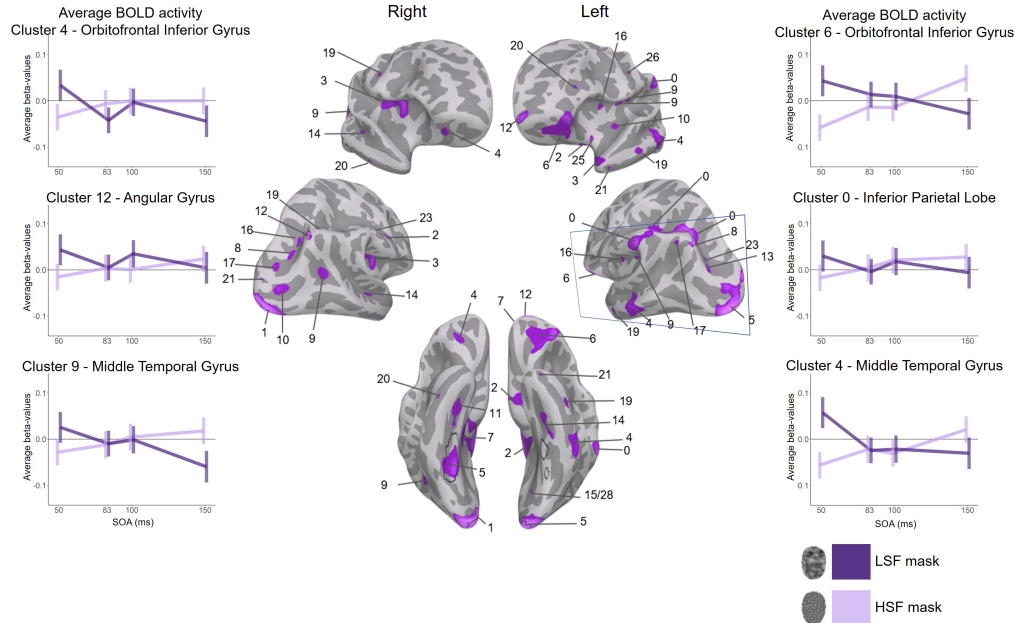


FIGURE 2.6: Whole-volume group-level analysis interaction map projected on an inflated brain surface. The thin box represents the brain volume that was scanned. Here we show the voxels with a significant SOA by SF masking interaction effect in the scanned volume. Purple highlighted areas show the difference in the fitted slope (HSF minus LSF) such that positive values support the coarse-to-fine signature (LSF masking more disruptive when presented early than late = more negative slope and vice versa for HSF masking). The map was FDR corrected for multiple comparisons, and no surviving vertices showed a positive value for the slope difference (i.e. we did not observe any significant 'fine-to-coarse signature'). For comparison with our ROI analyses, the black outlines indicate the location of the functionally defined V1 and FFA - collapsed over all subjects. The averaged beta-values in the plots are normalised.

TABLE 2.4: Whole-volume group-level analysis. Regions showing the interaction effect of SF masking with SOA (coarse-to-fine signature). For each cluster, the hemisphere (R = right hemisphere; L = left hemisphere), Brodmann's area (BA), number of voxels (N), and MNI152 coordinates (x, y, z). Based on the median voxel in each cluster of voxels.

Cluster nr		hemi	BA	N voxels	x	y	z
<u>Frontal areas</u>							
4	Frontal Inf Orb	R	47	131	29	35	-8
6	Frontal Inf Orb	L	47	1152	-35	37	-10
20	Frontal Inf Tri	L	46	20	-35	34	13
7	Frontal Med Orb	L	10	301	-9	54	-7
12	Frontal Sup Orb	L	10	134	-16	57	-8
2	Postcentral	R	4	145	48	-6	27
23	SupraMarginal	R	40	4	51	-23	36
26	Postcentral	L	4	45	-48	-14	32
<u>Temporal areas</u>							
5	Fusiform	R	37	650	40	-52	-15
7			36	396	35	-30	-17
11			20	315	41	-15	-23
28	Fusiform	L	19	2	-26	-75	-3
14	Temporal Mid	R	21	104	45	-26	-7
4	Temporal Mid	L	21	856	-57	-38	-9
19			21	111	-59	-13	-21
9	Temporal Mid	R	39	263	40	-57	16
20	Temporal Inf	R	21	14	51	-12	-26
14	Temporal Inf	L	20	203	-45	-25	-21
21			38	13	-42	3	-33
3	Temporal Pole Mid	L	38	410	-35	11	-31
<u>Occipital areas</u>							
1	Lingual	R	19	2012	18	-91	-5
8	Occipital Mid	R	39	85	29	-69	23
10			18	169	29	-83	8
23	Occipital Mid	L	19	28	-29	-72	20
13			19	74	-33	-81	18
8			39	54	-35	-75	38
16	Occipital Sup	R	47	77	29	-60	31
17			19	84	23	-83	19
21			18	12	22	-94	16
5	Occipital Inf	L	18	1875	-20	-95	-4
15	Calcarine	R	17	25	9	-91	11
1	Calcarine	L	17	259	-9	-88	10
27	Lingual	L	19	12	-15	-51	-2
15			19	13	-25	-72	-2
<u>Parietal areas</u>							
0	Parietal Inf	L	39	3466	-40	-49	40
3	Insula	R	13	1636	35	-8	17
16	Insula	L	13	75	-33	0	12
25			13	44	-34	7	-11
9			1	286	-34	-24	22
19	Angular	R	39	10	41	-58	44
12			39	174	37	-67	43
17	Angular	L	39	42	-39	-64	38
29	Precuneus	L	31	6	-17	-41	47

Psychophysiological Interaction

A premise of coarse-to-fine theories is that visual perception involves recurrent interactions between V1 and higher level regions in dorsal, frontal, and ventral brain regions. FFA has been shown to be particularly sensitive to any disruption of the natural statistics of the upright human face (e.g., Schiltz and Rossion, 2006; Yue et al., 2013), suggesting it may contain a template of the human face guiding coarse-to-fine integration. We thus tracked the influence of SF masking on the PPI connectivity between V1 and FFA.

The final LMM after pruning, however, only included stimulus type as fixed effect. There was just a small proportion of residual variability accounted for by the global differences between subjects (.004%). We found no effect of SOA nor of SF masking. The connectivity between V1 and FFA was stronger for naturally contrasted faces than for negated faces ($F = 33.69$, $p < .001$, see Figure 2.7). This effect cannot be attributed to differences in response amplitude between natural and negated faces, and suggests that the low and high tiers of the ventral pathway cooperate more strongly when the stimulus matches the internal template representation of the human face (i.e., face with a natural contrast polarity).

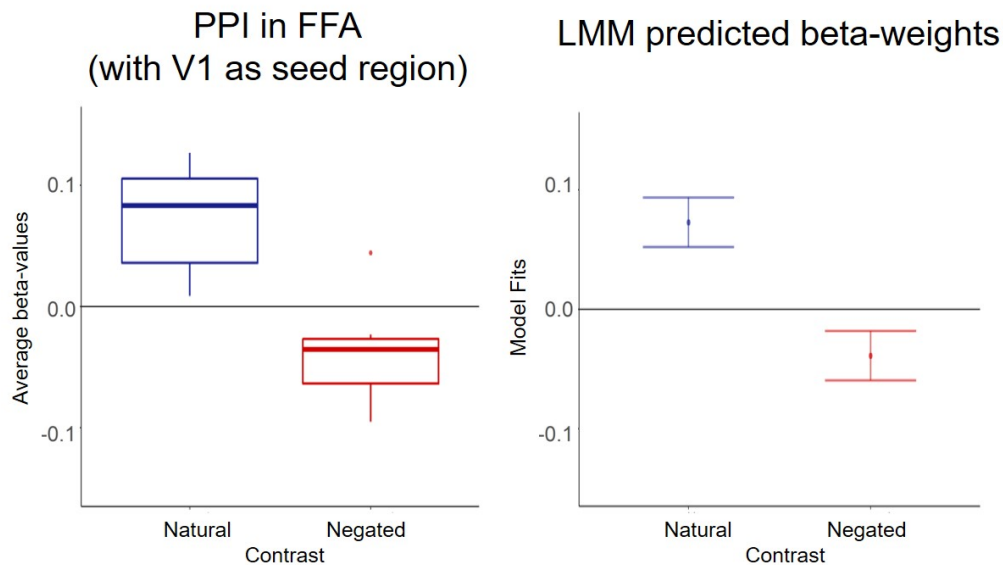


FIGURE 2.7: Results of the PPI analysis in the FFA with V1 as seed region. Left. Boxplots of the main effect of stimulus type on the PPI connectivity: upper and lower hinges correspond to the first and third quartile, whiskers represent a 95% within-subject confidence interval (Loftus & Masson, 1994) and the single dot represents an outlier. Right. Boxplots showing the model fits of the main effect of stimulus type (with 95% confidence interval) of a model with only stimulus type as main effect. The results reflect the differential connectivity patterns between V1 and FFA when processing intact face stimuli compared to their scrambled counterparts.

Discussion

The notion that recurrent processing contributes to the refinement of the coarse representations initially extracted by the visual system (Kok & de Lange, 2014) is influential in vision and computer sciences (Bar, 2004; Bullier, 2001; Kreiman & Serre, 2020; Mohsenzadeh et al., 2018). In this framework, it has been proposed that V1 plays the role of an 'active blackboard' for the coarse-to-fine integration of visual input (Bullier, 2001; Deco & Lee, 2004; Roland, 2010). We reasoned that if coarse-to-fine processing is implemented in V1 as an efficient way to represent visual information, its disruption should have measurable costs, and result in suboptimal representations in this region. We disrupted coarse-to-fine integration via backward SF masking to prevent the contribution of coarse and fine spatial scales over the course of human face recurrent processing (Bacon-Macé et al., 2005; Breitmeyer, 1984; Fahrenfort et al., 2007; Grill-Spector et al., 2000; Kovács et al., 1995; Lamme et al., 2002; Rolls & Tovee, 1994; Rolls et al., 1999). Disrupting LSF processing early was predicted to disrupt top-down LSF-driven signals, prevent the formation of a sparse representation, and therefore increase V1's response to the face stimulus. Our results support this prediction. Since LSFs contribute less to processing over time, disrupting LSF late was expected to be less disruptive, which was also supported by our results. The opposite trend was observed for HSF masking in line with the notion that fine cues, i.e. HSFs, are processed at a later stage, contributing progressively more to the sharpening of visual representations. In line with this notion, we found that the interference caused by HSF masking on V1 response increased over time. Our finding that the V1 response to strictly identical sets of full-spectrum face images follows opposite trends depending on the masked SF is compelling evidence of the complementary contribution of the coarse and fine scales to the build-up of progressively finer visual representations in V1 (e.g., Bar, 2004; Bullier, 2001; Hegdé, 2008). By retaining a core aspect of naturalistic vision; i.e. full range of aligned SF, our work goes beyond past evidence of a coarse-to-fine precedence, and yields novel insights on a fundamental tenet of coarse-to-fine models, namely that the integration of visual information operates in a coarse-to-fine manner in V1.

Current models of vision defend the importance of recurrent feedforward and feedback interactions between low and high processing levels (Ahissar & Hochstein, 2004; Bullier, 2001; Edelman & Gally, 2013; Keyser et al., 2001; Lamme & Roelfsema, 2000; Mumford, 1992). The current study builds upon this notion. We designed our experiment to minimise the possibility that the observed modulations of brain responses arise solely due to differences in feedforward processing. First, we employed identical full-spectrum stimuli across all conditions, which ensures that the SF, luminance and contrast of the stimuli were held constant across conditions. Second, by comparing the brain responses to intact and scrambled conditions, which were masked in strictly identical ways, our approach subtracted the direct influence of the mask on the BOLD response (e.g., Fahrenfort et al., 2007; Scholte et al., 2008) and controlled for the direct influence of the mask on the low-level mechanisms tuned to stimulus orientation/SF content.

Furthermore, our past MEG research indicates that phase scrambling does not influence initial feedforward responses in V1/V2 (Petras et al., 2021). This supports past suggestions that the intact-scrambled response differential analysed here is indicative of recurrent processing. Last but not least, we used backward masking to address the contribution of low and high SF over time. Past backward masking studies showed that the feedforward visual responses are well preserved (Enns & Di Lollo, 2000; Fahrenfort et al., 2007; Lamme et al., 2002; Ro et al., 2003; Rolls et al., 1999). Consistent evidence indicates that backward masking disrupts recurrent processing selectively, even with a SOA as short as 16.667ms (e.g., Fahrenfort et al., 2007). In our study, SOAs ranged from 50 to 150ms, which makes feedforward accounts even less likely (e.g. Lamme, 1995; Roland, 2010; Scholte et al., 2008; Takagaki et al., 2008). Yet, despite the fact that we designed our experiment to specifically tap into recurrent processing, the present data on their own are not sufficient to fully exclude feedforward accounts; our conclusions therefore necessarily rely on past literature.

Despite fMRI having a temporal resolution of several seconds, past and present results confirm that this technique can be used to address time-sensitive questions if combined with time-sensitive methods such as masking (e.g. Goffaux et al., 2011; Green et al., 2005; Grill-Spector et al., 2000; Haynes et al., 2005). The backward masking of specific SF ranges allowed us to address when and how the coarse and fine visual cues contribute to the build-up of face representations in V1. The drawback of our approach is that it is limited to a few a priori SOAs in contrast to other techniques such as EEG and MEG that allow for a denser temporal sampling. Hansen et al. (2021) recently developed an EEG encoding-based mapping technique to map the variance of human visual ERPS onto the variance in location-specific SF content of a large dataset of natural scenes. They observed a coarse-to-fine shift in visual encoding over the first 150ms, which validates the time window used in the present experiment. The authors reported additional coarse-to-fine coding phases starting 180ms and 260ms after stimulus onset, suggesting that longer SOAs should be used in future studies. The top-down signal driving the progressive sharpening of V1 representations is presumably a coarse representation of the visual input stored in more anterior high-level visual regions, the large receptive fields of which enable a more global representation of the stimulus (e.g., T. S. Lee, 2015). This tenet is supported by past findings that the high-level ventral regions found to preferentially respond to faces or places attune to LSFs more rapidly than to HSF content of their preferred stimulus class (Goffaux et al., 2011; Kauffmann, Chauvin, Pichat, & Peyrin, 2015; Musel et al., 2014; Peyrin et al., 2006; Schyns & Oliva, 1994). Here, we show that, in full-spectrum viewing conditions, the disruption of coarse-to-fine integration modulates face processing in the human FFA similarly to V1. Indeed, despite the fact that the FFA diverges from V1 in its overall bigger response to intact faces than to scrambled faces and its steady increase with SOA (i.e., face visibility; e.g., Kay and Yeatman, 2016; Yue et al., 2013), the response in FFA increases more rapidly when disrupting HSF than LSF. The whole-volume results confirmed that, in the ventral pathway,

coarse-to-fine integration mainly involves V1 and FFA (for the stimuli used in our experiment). However, when localised at the individual level, we also found the OFA temporal profile to mirror the V1 coarse-to-fine dynamics (see Supplementary material 6). Considering the closer proximity of OFA to V1 and the existence of denser connections between V1 and OFA than FFA (Finzi et al., 2021), it may come as a surprise that this region did not show up in the whole-volume analysis. Consistent evidence has nonetheless indicated that the OFA encodes a finer scale of face information than FFA (Goffaux et al., 2011, 2012; Rossion, Caldara, et al., 2003; Schiltz & Rossion, 2006), suggesting that, in contrast to FFA which would activate from the earliest LSF-driven stages, OFA may mainly contribute to the later stages of face representation refinement (Gentile et al., 2017; Rossion, Caldara, et al., 2003).

We used contrast negation to address whether the observed coarse-to-fine dynamics involve a priori knowledge of the natural contrast polarity of the human face (e.g., Bar, 2004; Mumford, 1992). We replicate past findings that negation substantially hampers the accumulation of sensory input necessary for the emergence of high-level category-selective neural responses (Liu-Shuang et al., 2022; see also Bex and Makous, 2002). However, the coarse-to-fine dynamics observed in V1 (and FFA) were not influenced by negation. This suggests that the (coarse-to-fine) sharpening of face representation in V1 is guided by high-level neural populations that are sensitive to phase-aligned content irrespective of whether the input matches the natural contrast statistics of an internally stored category. Still, the significant PPI connectivity between V1 and FFA for naturally contrasted faces indicates that these regions interact during the formation of face-specialised representations. Considering V1’s relative immunity to negation (George et al., 1999; Yue et al., 2013 but see Bex and Makous, 2002), the modulation of the V1/FFA connectivity by negation is likely to result from a reduced top-down influence of FFA on primary visual processing when the stimulus fails to correspond to the internal template representation of a face. Yet, the fact that negation did not influence the coarse-to-fine dynamics in V1 suggests these are driven by other regions than the FFA, as supported by the whole-volume results (see below). Future investigation is however needed to support this interpretation of the data since our study may simply be underpowered to reveal the influence of contrast negation on the coarse-to-fine dynamics of V1 and on its connectivity to FFA.

Besides the FFA, our whole-volume analyses showed a coarse-to-fine signature in orbitofrontal and dorsal (parietal) regions. This agrees with growing evidence that object recognition crucially relies on the involvement of dorsal and frontal regions (Ayzenberg & Behrmann, 2022; Bar, 2004; Kay & Yeatman, 2016; Zachariou et al., 2017). These regions are thought to rapidly generate a global and coarse representation of incoming input, i.e. an initial 'guess' (Ayzenberg & Behrmann, 2022; Ayzenberg et al., 2023; Bar et al., 2006; Fintzi & Mahon, 2014; Kauffmann, Bourgin, et al., 2015; Kauffmann, Chauvin, Pichat, & Peyrin, 2015; Kveraga et al., 2007; Peyrin et al., 2010; Zachariou et al., 2017). This is supported by past findings that the dorsal pathway represents shape before - and sends information to - the ventral pathway (Ayzenberg & Behrmann, 2022; Van Dromme et al., 2016). During human face processing, in particular, the posterior parietal cortex has been proposed to represent the global configuration of facial features and convey it to the face-selective ventral regions (Zachariou et al., 2017). In line with the absence of a negation effect upon V1's coarse-to-fine dynamics, our results suggest that the coarse structure guiding the representation sharpening in V1 mainly comes from dorsal and orbitofrontal non face-selective populations. Altogether our findings therefore challenge the hypothesis that coarse-to-fine integration occurs primarily in inferotemporal cortex (e.g., Bar, 2004; Kveraga et al., 2007; Wyatte et al., 2014). They instead favour the notion that the coarse representation activated early in dorsal and orbitofrontal cortices is integrated with the representation of the fine details in V1. In this context, the inferotemporal cortex may not play an active role in coarse-to-fine integration but would encode the output signals delivered by the dorso-frontal-V1 coarse-to-fine network (see Deco and Lee, 2004 for a similar proposal).

In the present study, the utilisation of an orthogonal task was deemed necessary, given the impracticability of performing a face-related task in the context of phase-scrambled conditions. The collected behavioural data indicates that subjects were equally able to perceive the rare colour changes of the stimulus. However, whether and how the disruption of coarse-to-fine processing influences face perception per se remains elusive. Faces can be categorised at diverse levels of precision (from detection to individual identity recognition). Evidence from past Rapid Serial Visual Presentation (RSVP) studies, in which the perception of each stimulus is masked both by the preceding (forward masking) and following stimulus (backward masking), set a lower bound for the duration necessary to detect, categorise or recognise a face at the individual identity level. Using this approach, Retter et al. (2020) showed that human observers are able to perform face categorisation (i.e., detect the occurrences of face stimuli in a wide diversity of natural images with a group-level accuracy of 68%) for presentation durations as short as 50ms. This suggests that our subjects had no difficulty to differentiating backward-masked intact from scrambled faces stimuli even at the shortest SOAs. Whether they were able to access the individual identity of the intact stimuli is another question. Only a few (RSVP) studies have addressed the timing of face individuation. They suggest that 50ms of face viewing is the

minimum stimulus duration to access individual identity cues; identity recognition accuracy is relatively low at such short presentation duration but increases steeply within the first 100ms of stimulus duration (e.g. Tanskanen et al., 2007). Our use of backward masking alone resulted in less disruption than is seen in RSVP studies (Keysers & Perrett, 2002). Our preliminary findings using a behavioural backward SF masking paradigm similar to the one utilised here indeed suggest that observers are able to recognise the identity of familiar face celebrities above chance even at the shortest SOAs (50ms). Identity recognition is expected to improve as a function of SOA, and more so when masking HSF than the LSF of the face stimulus in line with coarse-to-fine hypotheses.

The finding that masking of specific SFs differentially impacts V1 over processing course indicates that the contribution of this region to visual processing is sustained, which challenges the still influential assumption that V1's role is limited to an early filtering of visual information to downstream regions (Kravitz et al., 2013). Our findings favour the proposal that V1 integrates incoming input with intrinsic and top-down orbitofrontal and dorsal influences. They add to the body of evidence suggesting that V1 is involved in higher levels of visual function than is commonly assumed (Alink et al., 2010; Astorga et al., 2022; Ayzenshtat et al., 2012; Chadwick et al., 2023; Kazandjian et al., 2021; Koivisto et al., 2011; Kok & de Lange, 2014; Lamme, 1995; Lamme et al., 1998; Muckli, 2010; Noudoost & Moore, 2011; Rademaker et al., 2019; Roland et al., 2006; Super et al., 2001; Zipser et al., 1996). Because the retinotopic format of V1 coding is shared across processing levels, it offers a common platform for bottom-up and top-down communication (Deco & Lee, 2004; Kravitz et al., 2013; Mumford, 1992). Yet, the finding that V1 is recruited for the coarse-to-fine processing of the human face is compelling. Indeed, human face perception is well documented for its reliance on high-level visual specialised neural populations (e.g., M. Bernstein et al., 2018; Kanwisher et al., 1997; Rossion, Schiltz, and Crommelinck, 2003; Tsao et al., 2006). Consequently, the potential contribution of V1 to human face processing has been largely overlooked (e.g., recent Collins and Olson, 2014; Fox et al., 2009; Gobbini and Haxby, 2007; Ishai, 2008; with some notable exceptions, e.g.: Grill-Spector et al., 2017; Petro et al., 2013). A fortiori, the coarse-to-fine dynamics found in V1 during face encoding indicate that the primary visual cortex acts as a generic blackboard for the recurrent refinement of representations in the human visual system (Bullier, 2001; Kok & de Lange, 2014; T. S. Lee et al., 1998). We thus believe that the present findings yield important insights for object recognition in general.

By combining backward masking with fMRI, we present evidence that the integration of visual information operates in a coarse-to-fine manner in V1. Interrupting the processing of coarse input structure disrupted V1 activity most severely in the earliest time window, and decreased in influence over time; the opposite was observed when masking HSF. Similar signatures of coarse-to-fine integration were found in fusiform, parietal and frontal regions. The robust coarse-to-fine dynamics observed in V1 indicate its active involvement in the progressive sharpening of visual representations in the human brain. The fact that V1 response dynamics to strictly identical stimulus sets differed depending on the masked scale adds to growing evidence that the role of V1 goes beyond the early processing and passive transmission of visual information to the rest of the brain. It instead indicates that V1 may yield a 'spatially registered common forum' or 'blackboard' where top-down inferences are integrated with incoming visual signals through recurrent interaction with high-level regions located in the inferotemporal, dorsal and frontal regions (Bullier, 2001; Deco & Lee, 2004; Roland, 2010).

Supplementary Material

Supplementary Material 1: Functional localiser stimuli

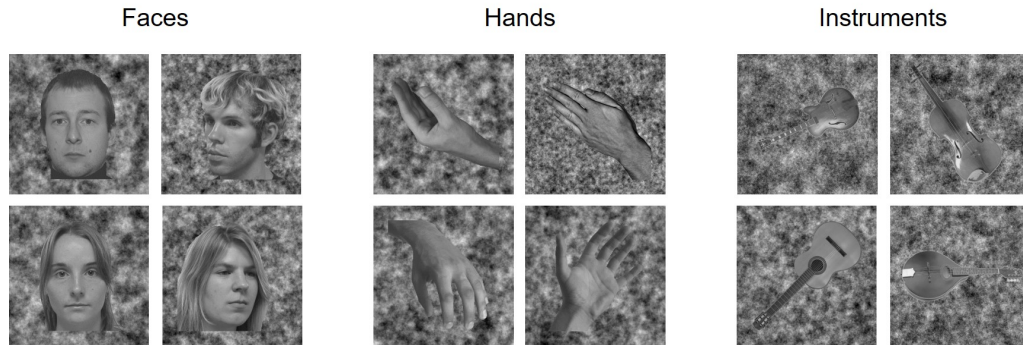


FIGURE 2.8: Stimuli as used in the functional localiser experiment. Stimuli were adapted from the fLoc functional localiser package (Stigliani et al., 2015). The background for the stimuli were created using the iterative phase scrambling method as described in the method section to ensure a uniform distribution of SF and orientation content over the whole image.

Supplementary Material 2: Different V1 delineation

In our study, we delineated V1 in each individual subject on their surface vertex based on a thresholding approach (Muckli et al., 2015; F. W. Smith & Muckli, 2010). We opted for this surface-based method to ensure the selection of a continuous region of the cortical surface and to avoid selecting voxels outside of this region (Gattass et al., 2005) due to the folding of the cortex (Dale et al., 1999). A thresholding approach in volume space could also have been an alternative option, we conducted additional analyses to examine the replicability of our findings using this approach.

Specifically, we identified voxels that exhibited a significantly stronger response to the face-containing stimulus region compared to its background by contrasting the responses to flickering checkerboard conditions in the block at the end of each run, using the [stimulus - background] contrast. We selected clusters of significant voxels ($t > 1.65$) in the early visual cortex from this contrast and subsequently integrated this cluster with the individual V1 label obtained from *Freesurfer*. Next, we selected the overlapping voxels as the V1 ROI for the current analysis. Notably, this more conservative approach, which did not involve data smoothing, resulted in ROIs that were 22.3 (SD = 10.3) percent of the size of the ROIs used in the main analysis. Despite this more stringent approach, we replicated the main results, providing compelling evidence for the robustness of our findings.

First, and to ascertain the negative nature of the intact-scrambled response difference under less constrained conditions, we conducted a comparison of the responses to intact and phase-scrambled faces from the functional localiser in V1. We found that the intact-scrambled response difference ($M = -.08$, $SD = .14$) was significantly lower than zero ($t(31) = -2.30$, $p = 0.036$). With main analysis, we observed the distinct "signature" indicative of coarse-to-fine processing in V1. Specifically, we found that early masking of LSFs led to heightened V1 responses to face stimuli, as evidenced by a positive intact-scrambled difference. The disruptive effect of LSF masking decreased over time. In contrast, we observed a reverse tendency for HSF masking, as the contribution of HSF processing to perception is hypothesised to increase with longer processing time.

Following a systematic removal of non-significant higher-order interaction terms, as determined by a *type III ANOVA*, our LMM included a two-way interaction effect terms of SF masking (see Supplementary Table 2.5). As no hemispheric differences were observed, we collapsed the data over both hemispheres in the LMM to enhance statistical power. The following final model was used to analyse the newly defined ROI data ($BIC = 117266.3$):

$$BetaValue \sim 1 + SOA * SFmask + (1|subject)$$

The proportion of residual variability accounted for by global differences between subjects was negligible (0.001%). Notably, we found a significant interaction effect between SOA and SF masking. Supplementary Figure 2.9 displays the average BOLD activity and the LMM predictions from the full model, incorporating all terms for illustrative purposes, despite not all of them being significant.

In both naturally contrasted and negated images, we observed a distinct cross-like pattern in the intact-scrambled response differential within V1. Specifically, when masking LSFs and collapsing over stimulus type, we observed a negative slope as a function of SOA ($\beta = -.12$, $CI = .14$), whereas masking HSFs yielded a positive slope ($\beta = .18$, $CI = .14$). Importantly, these slopes were found to be significantly different from each other ($z = -6.022$, $p < .001$). The estimated marginal means of the model and pairwise comparisons between LSF and HSF masked conditions are presented in Supplementary Table 2.6. Notably, at 50ms, the V1 response differential was significantly larger when LSFs were disrupted compared to HSFs. However, by 150ms, this trend was reversed, with disruption of HSFs leading to a significantly larger V1 response compared to disruption of LSFs (see Supplementary Table 2.6).

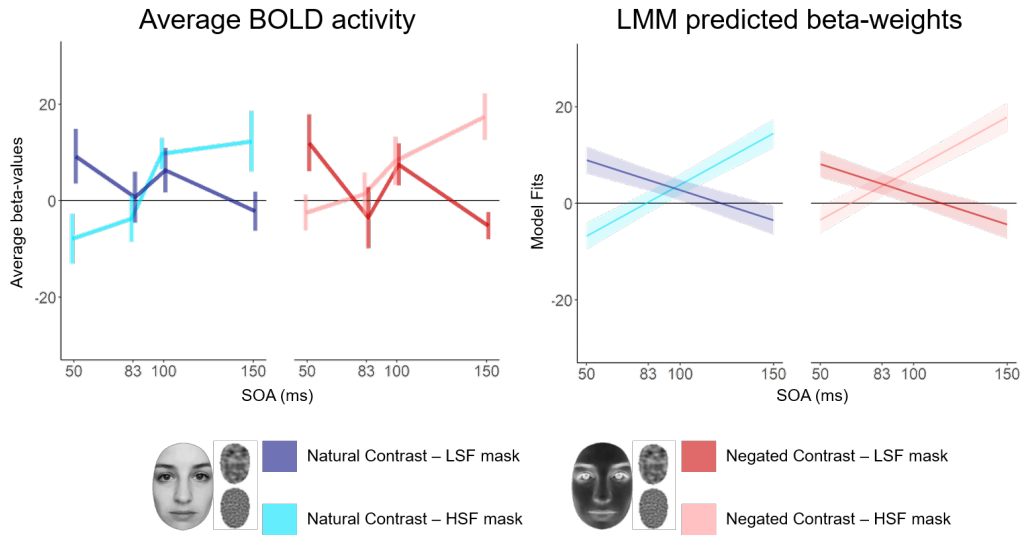


FIGURE 2.9: Intact-scrambled response differential for a differently defined V1 (i.e. thresholding approach in volume space instead of delineation on the surface vertex). Plots in the left column represent the group averaged beta-values including a 95% within-subjects confidence interval as described in Loftus and Masson (1994). The LMM fits shown on the right plots are from a full model including all interaction effects which was ran just for illustrative purposes. There was a significant interaction effect of SOA with SF masking.

TABLE 2.5: Type III ANOVA - for newly defined ROI data in V1

	<i>hemi</i>	<i>sum Sq</i>	<i>NumDF</i>	<i>DenDF</i>	<i>F</i>	<i>p</i>	
SOA	11739	11739	1	9869.6	1.4275	.232	
SFmask	226578	226578	1	9869.6	27.5526	<.001	***
SOA:Sfmask	298192	298192	1	9869.6	36.2611	<.001	***

TABLE 2.6: Model predictions (estimated marginal means - emmeans) and pairwise comparisons on predicted variables between LSF and HSF masked conditions for different SOAs in the newly defined V1 ROI.

<i>SOA</i>	<i>LSF mask</i> <i>emmean (SE)</i>	<i>HSF mask</i> <i>emmean (SE)</i>	<i>Estimate</i> <i>(SE)</i>	<i>Test statistic (z/t)</i>	<i>p</i>	
50	8.30 (2.23)	-3.62 (2.23)	11.9 (2.94)	4.05	<.001	***
83	4.28 (1.59)	2.39 (1.59)	1.89 (1.93)	0.975	.329	
100	2.21 (1.53)	5.49 (1.53)	-3.28 (1.84)	-1.79	.074	.
150	-3.89 (2.46)	14.60 (4.46)	-18.5 (3.29)	-5.62	<.001	***

TABLE 2.7: Newly defined V1 ROI Fixed effects

	<i>Estimate</i>	<i>SE</i>	<i>df</i>	<i>T</i>	<i>p</i>	
(Intercept)	0.83113	2.70828	364.293	0.307	.759	
SOA	0.03017	0.02525	9869.598	1.195	.232	
SFmask	13.56453	2.58418	9869.598	5.249	<.001	***
SOA:Sfmask	-0.15206	0.02525	9869.598	-6.022	<.001	***

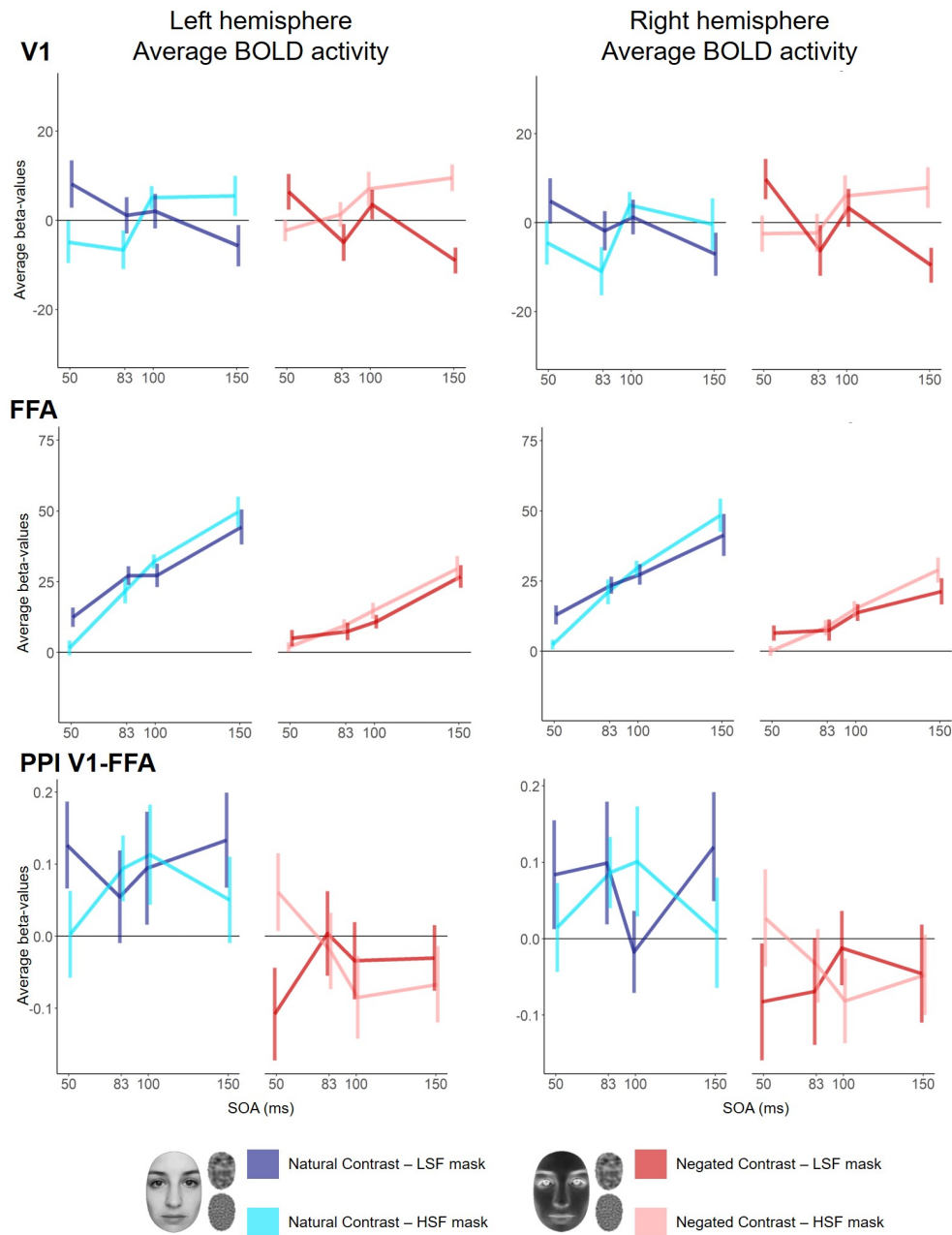
Supplementary Material 3: Plots of beta values from GLM

FIGURE 2.10: Intact-scrambled response differences for V1, FFA and for the connectivity analysis between these regions, for both left (left column) and right (right column) hemispheres. The group averaged beta-values include a 95% within-subjects confidence interval as described in Loftus and Masson (1994). Data is just for illustration, since no hemispherical differences were found in the analyses. Cartoon faces are used in replacement of photographs in the figure as journal regulations prohibit the use of identifiable faces.

Supplementary Material 4: Linear Mixed effect Models: step-wise variable selection

TABLE 2.8: Step-by-step overview of the stepwise variable selection procedure following backward elimination (Heinze et al., 2018). The data is either accuracy of the behavioural task, ROI data (V1, V2, FFA or OFA) or the PPI between V1 and FFA. The first model is the most complex model containing all interaction terms, next the non-significant (interaction) terms were stepwise eliminated from the model.

Model number	Model	BIC
Accuracy		
1	Accuracy $\sim 1 + \text{SOA} * \text{SFmask} * \text{StimulusType} + (1 \text{subject})$	-1298.851
2	Accuracy $\sim 1 + \text{SOA} * \text{SFmask} + \text{SOA} * \text{StimulusType} + \text{SFmask} * \text{StimulusType} + (1 \text{subject})$	-1343.635
3	Accuracy $\sim 1 + \text{SOA} + \text{SFmask} + \text{StimulusType} + (1 \text{subject})$	-1443.545
V1		
1	BetaValue $\sim 1 + \text{SOA} * \text{SFmask} * \text{StimulusType} * \text{Hemisphere} + (1 \text{subject})$	118743.7
2	BetaValue $\sim 1 + \text{SOA} * \text{SFmask} * \text{StimulusType} + \text{SOA} * \text{SFmask} * \text{Hemisphere} + \text{SOA} * \text{StimulusType} * \text{Hemisphere} + \text{SFmask} * \text{StimulusType} * \text{Hemisphere} + (1 \text{subject})$	118729
3	BetaValue $\sim 1 + \text{SOA} * \text{SFmask} + \text{SOA} * \text{StimulusType} + \text{SFmask} * \text{StimulusType} + \text{SOA} * \text{Hemisphere} + \text{SFmask} * \text{Hemisphere} + \text{StimulusType} * \text{Hemisphere} + (1 \text{subject})$	118676.6
4	BetaValue $\sim 1 + \text{SOA} * \text{SFmask} + \text{SFmask} * \text{StimulusType} + (1 \text{subject})$	118624.6
V2		
1	BetaValue $\sim 1 + \text{SOA} * \text{SFmask} * \text{StimulusType} * \text{Hemisphere} + (1 \text{subject})$	114363.5
2	BetaValue $\sim 1 + \text{SOA} * \text{SFmask} * \text{StimulusType} + \text{SOA} * \text{SFmask} * \text{Hemisphere} + \text{SOA} * \text{StimulusType} * \text{Hemisphere} + \text{SFmask} * \text{StimulusType} * \text{Hemisphere} + (1 \text{subject})$	114348.1
3	BetaValue $\sim 1 + \text{SOA} * \text{SFmask} + \text{SOA} * \text{StimulusType} + \text{SFmask} * \text{StimulusType} + \text{SOA} * \text{Hemisphere} + \text{SFmask} * \text{Hemisphere} + \text{StimulusType} * \text{Hemisphere} + (1 \text{subject})$	114294.2
4	BetaValue $\sim 1 + \text{SOA} * \text{SFmask} + \text{SFmask} * \text{StimulusType} + (1 \text{subject})$	114240.2
FFA		
1	BetaValue $\sim 1 + \text{SOA} * \text{SFmask} * \text{StimulusType} * \text{Hemisphere} + (1 \text{subject})$	107593.5
2	BetaValue $\sim 1 + \text{SOA} * \text{SFmask} * \text{StimulusType} + \text{SOA} * \text{SFmask} * \text{Hemisphere} + \text{SOA} * \text{StimulusType} * \text{Hemisphere} + \text{SFmask} * \text{StimulusType} * \text{Hemisphere} + (1 \text{subject})$	107577.8
3	BetaValue $\sim 1 + \text{SOA} * \text{SFmask} + \text{SOA} * \text{StimulusType} + \text{SFmask} * \text{StimulusType} + \text{SOA} * \text{Hemisphere} + \text{SFmask} * \text{Hemisphere} + \text{StimulusType} * \text{Hemisphere} + (1 \text{subject})$	107523.6
4	BetaValue $\sim 1 + \text{SOA} * \text{SFmask} + \text{SOA} * \text{StimulusType} + (1 \text{subject})$	107475
OFA		
1	BetaValue $\sim 1 + \text{SOA} * \text{SFmask} * \text{StimulusType} * \text{Hemisphere} + (1 \text{subject})$	113935.8
2	BetaValue $\sim 1 + \text{SOA} * \text{SFmask} * \text{StimulusType} + \text{SOA} * \text{SFmask} * \text{Hemisphere} + \text{SOA} * \text{StimulusType} * \text{Hemisphere} + \text{SFmask} * \text{StimulusType} * \text{Hemisphere} + (1 \text{subject})$	113920.3
3	BetaValue $\sim 1 + \text{SOA} * \text{SFmask} + \text{SOA} * \text{StimulusType} + \text{SFmask} * \text{StimulusType} + \text{SOA} * \text{Hemisphere} + \text{SFmask} * \text{Hemisphere} + \text{StimulusType} * \text{Hemisphere} + (1 \text{subject})$	113866.5
4	BetaValue $\sim 1 + \text{SOA} * \text{SFmask} + \text{SOA} * \text{StimulusType} + (1 \text{subject})$	113824.1
PPI (V1-FFA)		
1	BetaValue $\sim 1 + \text{SOA} * \text{SFmask} * \text{StimulusType} * \text{Hemisphere} + (1 \text{subject})$	28709.61
2	BetaValue $\sim 1 + \text{SOA} * \text{SFmask} * \text{StimulusType} + \text{SOA} * \text{SFmask} * \text{Hemisphere} + \text{SOA} * \text{StimulusType} * \text{Hemisphere} + \text{SFmask} * \text{StimulusType} * \text{Hemisphere} + (1 \text{subject})$	28692.92
3	BetaValue $\sim 1 + \text{SFmask} * \text{StimulusType} + \text{SOA} * \text{SFmask} + \text{SOA} * \text{StimulusType} + \text{SOA} * \text{Hemisphere} + \text{SFmask} * \text{Hemisphere} + \text{StimulusType} * \text{Hemisphere} + (1 \text{subject})$	28623.33
4	BetaValue $\sim 1 + \text{SOA} + \text{SFmask} + \text{Hemisphere} + \text{StimulusType} + (1 \text{subject})$	28529.54
5	BetaValue $\sim 1 + \text{StimulusType} + (1 \text{subject})$	28482.8

Supplementary Material 5: Tables with fixed effects

TABLE 2.9: Accuracy Fixed effects

	<i>Estimate</i>	<i>SE</i>	<i>df</i>	<i>T</i>	<i>p</i>	
(Intercept)	0.97260	0.00487	225.6	199.824	<.001	***
SOA	-0.00002	0.00005	364	-0.542	.588	
SFmask	-0.00181	0.00164	364	-1.103	.271	
StimulusType[natural]	0.00215	0.00232	364	0.928	.354	
StimulusType[negated]	0.00181	0.00232	364	0.778	.437	

TABLE 2.10: V1 Fixed effects

	<i>Estimate</i>	<i>SE</i>	<i>df</i>	<i>T</i>	<i>p</i>	
(Intercept)	1.724	2.392	300.3	0.721	.472	
SOA	-0.01528	0.02204	10190	-0.693	.488	
SFmask	11.03	2.256	10190	4.892	<.001	***
StimulusType	0.8803	0.796	10190	1.106	.269	
SOA:Sfmask	-0.1204	0.02204	10190	-5.462	<.001	***
SFmask:StimulusType	-1.466	0.796	10190	-1.842	.066	.

TABLE 2.11: FFA Fixed effects

	<i>Estimate</i>	<i>SE</i>	<i>df</i>	<i>T</i>	<i>p</i>	
(Intercept)	-9.929	2.368	27.77	-4.193	<.001	***
SOA	0.3092	0.01274	10190	24.279	<.001	***
SFmask	6.397	1.303	10190	4.908	<.001	***
StimulusType	0.03893	1.303	10190	0.03	.976	
SOA:Sfmask	-0.06636	0.01274	10190	-5.21	<.001	***
SOA:StimulusType	-0.07075	0.01274	10190	-5.555	<.001	***

TABLE 2.12: PPI Fixed effects

	<i>Estimate</i>	<i>SE</i>	<i>df</i>	<i>T</i>	<i>p</i>	
(Intercept)	0.01682	0.01831	15.02	0.919	.373	
StimulusType	-0.05591	0.009632	10190	-5.804	<.001	***

TABLE 2.13: V2 Fixed effects

	<i>Estimate</i>	<i>SE</i>	<i>df</i>	<i>T</i>	<i>p</i>	
(Intercept)	0.5065	1.912	0.03428	0.265	.791	
SOA	9.976	1.820	1.019e+04	5.483	<.001	***
SFmask	0.07532	0.621	1.019e+04	0.117	.907	
StimulusType	0.02686	0.01778	1.019e+04	1.511	.13	
SOA:Sfmask	-1.368	0.6421	1.019e+04	-2.130	.033	*
SFmask:StimulusType	-0.1022	0.01778	1.019e+04	-5.747	<.001	***

TABLE 2.14: OFA Fixed effects

	<i>Estimate</i>	<i>SE</i>	<i>df</i>	<i>T</i>	<i>p</i>	
(Intercept)	-13.73	3.9060	22.39	-3.516	.002	**
SOA	0.461	0.0174	10190	26.528	<.001	***
SFmask	-3.722	1.7780	10190	-2.093	.036	*
StimulusType	7.920	1.7780	10190	4.453	<.001	***
SOA:Sfmask	-0.0493	0.0174	10190	-2.834	.005	**
SOA:StimulusType	-0.0660	0.0174	10190	-3.799	<.001	***

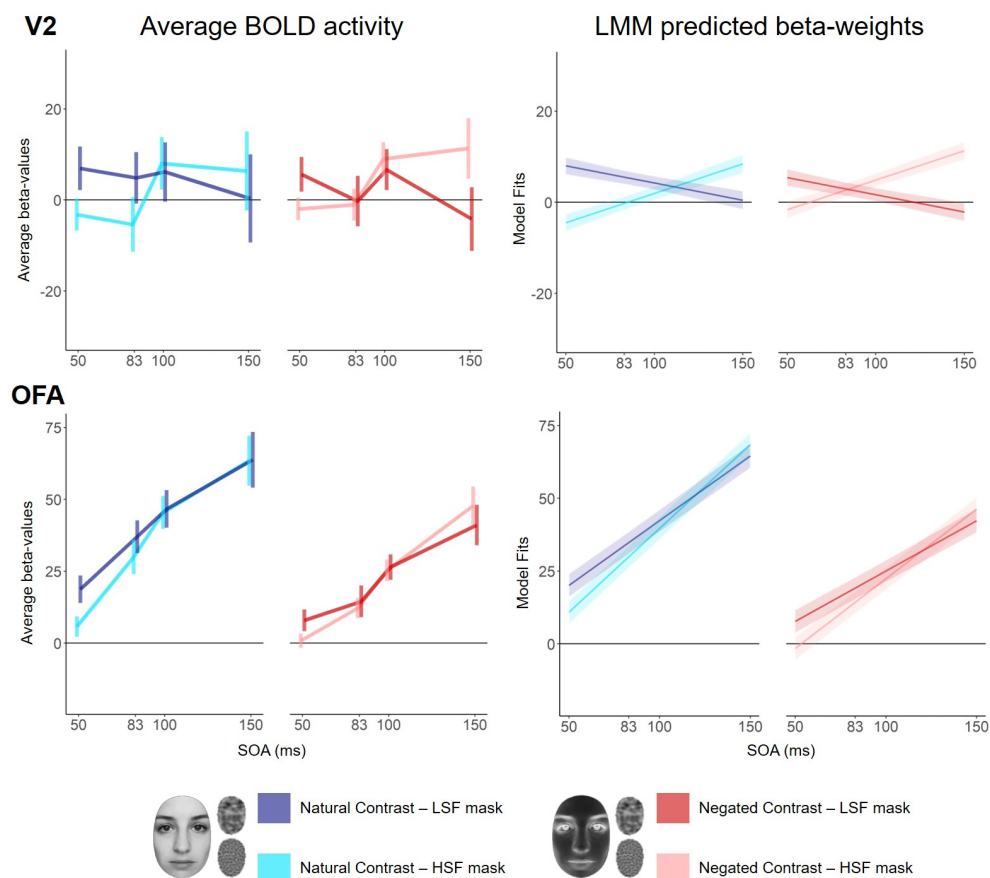
Supplementary Material 6: Extra figures: V2 and OFA

FIGURE 2.11: Intact-scrambled response differential for extra ROIs V2 (top; anatomically defined with Freesurfer label) and OFA (bottom; functionally defined with the functional localiser). Line plots in the left column represent the group averaged beta-values including a 95% within-subjects confidence interval (Loftus & Masson, 1994). The LMM predictions shown on the right plots are from a full model including all interaction effects with was ran just for illustrative purposes.

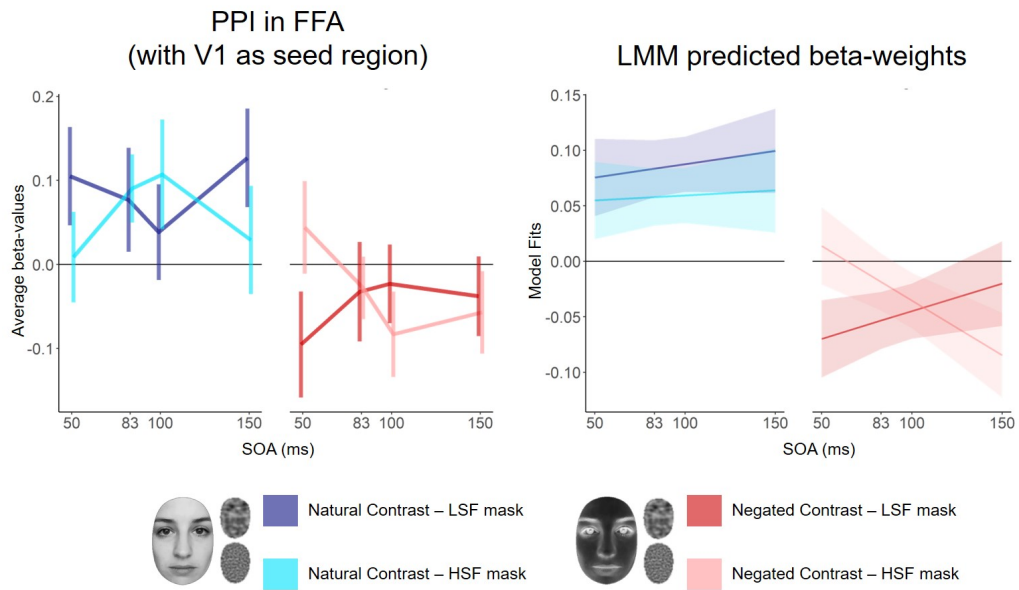


FIGURE 2.12: Results of the PPI analysis in the FFA with V1 as seed region. Left. Line plot shows PPI results averaged over all subjects including a 95% within-subjects confidence interval (Loftus & Masson, 1994). Right. Line plot with the model predictions of the full model including all interaction terms, since there were no significant three-way or two-way interactions (only a main effect of stimulus type), the plots are just for illustrative purposes.

Chapter 3

The role of coarse-to-fine integration in human face identity recognition

Jolien P. Schuurmans¹, Mrityika Dey¹, Valérie Goffaux^{1,2}

¹Psychological Sciences Research Institute, Université Catholique de Louvain, Louvain-la-Neuve, Belgium; ²Institute of Neuroscience, Université Catholique de Louvain, Louvain-la-Neuve, Belgium

Abstract

Natural input contains multiple SFs, encoded from the earliest stages of processing by the visual system. Spatial frequency (SF) processing is theorised to proceed in a coarse-to-fine manner, with low SFs (LSFs) processed initially, guiding the slower encoding of high SF (HSF) information. This was supported by our past neuroimaging study, addressing coarse-to-fine processing in the framework of face perception. Disrupting this coarse-to-fine integration of face stimuli using backward masking modulated responses in V1, ventral, dorsal and orbitofrontal regions. However, the behavioural implications of these findings, particularly in the context of familiar face recognition, remained unexplored. Here, we investigate the impact of disrupting coarse-to-fine processing on face identity recognition behaviour. We expected that LSF processing is fast and transient and less sensitive to backward masking while HSF processing is slower, more sustained and reliant on recurrent processes, therefore more masking-sensitive. To test this, we presented full-spectrum face images of familiar celebrities whose LSFs ($<11.36\text{cpf}$) or HSFs ($>11.36\text{cpf}$) were disrupted using backward masking at gradually increasing stimulus onset asynchronies (SOAs; 50, 75, 100, 150 and 300ms). We manipulated the signal-to-noise ratio (SNR) of the target face using a staircase method to assess the threshold SNR required to achieve 75% accuracy in all conditions. We expected to find two things. First, decreasing SNR thresholds as a function of SOA, since more processing time would increase the possibility to recognise identity. Secondly, we expected the decrease to be more pronounced for HSF masking compared to LSF masking, which would mirror the processing profile previously observed in the fusiform face area (FFA). While we did indeed find improved identity recognition performance as a function of SOA, we did not find any difference between SF masking conditions. The findings are inconclusive regarding the coarse-to-fine framework and are an indication

that our task may not have been sensitive enough to capture changes between conditions. They prompt future research to clarify the potential contribution of coarse-to-fine processing to familiar face recognition.

Introduction

Natural input contains information at multiple spatial frequencies (SFs) (see Figure 1). SF refers to the variation in intensity and spatial arrangement of visual information across different scales (Collin et al., 2004; DeValois et al., 1990; Graham, 1989; Westheimer, 2001). Not only are SFs encoded early, but they also exert influence on vision up to its higher processing stages, such as face recognition (e.g. C. H. Liu et al., 2000; Näsänen, 1999). Additionally, our ability to perceive and differentiate between individuals is aided by the presence SFs (e.g. Boutet et al., 2003).

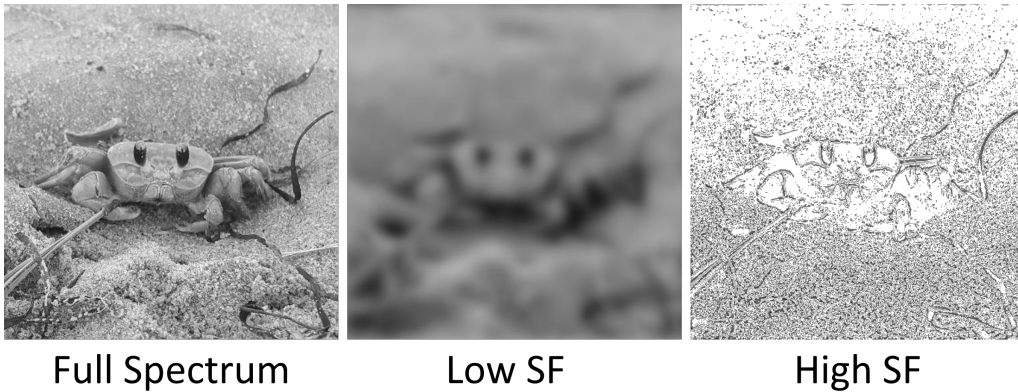


FIGURE 3.1: Depiction of SFs in a crab photograph. The left image is the original grayscale picture, incorporating all SF ranges (i.e., full-spectrum). The central image illustrates the same photo, but only with LSFs, generated with a low-pass Butterworth filter (SF cut-off: 11). The right image, processed with a high-pass Butterworth filter (SF cut-off: 20), reveals only HSFs from the initial photo.

Faces consist of a range of SF components that collectively contribute to the overall appearance and encoding of facial features. At the coarser scales, global face structure is captured, whereas finer scales delineate more specific facial characteristics, such as eye shape, lip contour, and skin texture (e.g. Goffaux et al., 2009; Goffaux and Rossion, 2006; Landi and Freiwald, 2017; C. H. Liu et al., 2000). Although a full range of SFs exist in faces, early studies reported that face identities can be recognised even at extremely low SFs (LSFs) (Harmon, 1973; Harmon & Julesz, 1973; Rubin & Siegel, 1984; Sinha, 2002; Yip & Sinha, 2002), highlighting the role of LSFs (around <5 cycles per face width; cpf) in driving the human brain's preferential response to face categories (Goffaux et al., 2003; Quek et al., 2018). The specific range that proves essential for the visual system depends on the demands of the task at hand (Schyns, 1998). More

intermediate SF ranges (8-16cpf) have been underscored for their significance in face identity recognition (e.g., Collin et al., 2004; Costen et al., 1994, 1996; Fiorentini et al., 1983; Gold et al., 1999; Kornowski and Petersik, 2003; C. H. Liu et al., 2000; Näsänen, 1999; Parker and Costen, 1999; Rotshtein et al., 2007), with an optimal mean of 11.1cpf for the most effective recognition performance (Costen et al., 1996). These findings collectively emphasise the critical role of mid-band SF bands in human face perception.

The past studies reviewed here, however, have focused on identifying the optimal SF range for processing face information by selectively filtering SF content in faces, essentially examining faces presented with only specific SF bands. Most studies overlook the fact that natural input consists of a wide spectrum of spatially aligned SFs. How exactly the visual system integrates input SFs is still a fundamental question in vision sciences. Several authors have proposed that SFs are integrated in a coarse-to-fine manner. It has been consistently observed that coarse/LSFs are processed before fine/high SFs (HSF) during the analysis of complex visual stimuli (Bar, 2004; Bullier, 2001; Hegdé, 2008; Kreiman & Serre, 2020; Watt, 1987). Several theoretical frameworks and models propose the importance of coarse-to-fine integration in visual perception. For instance, the model by Bullier (2001) suggests that LSF information is initially transmitted from the lateral geniculate nucleus (LGN) to primary visual regions (V1 and V2) and rapidly to the parietal cortex through the dorsal magnocellular pathway. The activation of the ventral pathway, which includes category-selective areas (i.e. Fusiform Face Area, FFA; Parahippocampal Place Area, PPA), occurs slightly later. This information is then fed back to V1 and V2, guiding the processing of HSFs via the slower and sustained parvocellular pathway. In this framework, HSF processing relies more heavily on recurrent interactions, both bottom-up (from primary to higher visual regions) and top-down (from higher to primary visual regions), compared to LSF processing. The resulting integration process refines visual representations, transitioning from coarse/LSF to fine-grained/HSF representations (Bullier, 2001).

Various studies have provided valuable insights into coarse-to-fine precedence. For example, Yan et al. (2022) investigated the recognition of familiar faces, by presenting different unfamiliar faces at a speed of 6Hz, with familiar celebrity faces every sixth face. The faces were presented in a temporal sequence of increasing (coarse-to-fine) or decreasing (fine-to-coarse) SF ranges. They found that recognition thresholds varied across different presentation conditions, with the coarse-to-fine sequence requiring fewer SF details for full recognition compared to the fine-to-coarse sequence. These findings suggest that spatial information in a coarse-to-fine sequence is integrated more efficiently. Moreover, Goffaux et al. (2011) utilised backward masking in a neuroimaging study to interrupt the processing of different SF ranges of face information at specific times. They demonstrated that high-level face specialised regions in the human visual cortex represent coarse scales of face stimuli before its finer content. Studies by Kauffmann, Bourgin, et al. (2015), Kauffmann et al. (2017), and Schyns and Oliva (1994) further demonstrated that the coarse-to-fine nature of natural scene

processing, with LSF information dominating perception during brief exposures, while HSF information becomes more influential with longer exposures, their findings indicate that coarse information precedes and can impact the processing of fine-grained information. These studies support the notion that LSF processing occurs early and transiently, while HSF processing is more sustained and probably more reliant on recurrent processes.

While coarse-to-fine precedence is well documented, empirical evidence for coarse-to-fine integration remains somewhat limited. The challenge lies in proving not only that coarse image information processes quicker than finer details but also that this initial processing directly influences the subsequent integration of HSF information. Addressing this question necessitates a more ecologically valid approach using full-spectrum visual stimuli that encompass all SF ranges aligned in space. Herein lies the significance of the studies by Petras et al. (2019, 2021), they used EEG and MEG to study this integration process in response to broadband face stimuli to address whether such coarse-to-fine integration results in a computational advantage as proposed by coarse-to-fine theories. In the EEG study, Petras et al. (2019) found a late decrease in HSF processing specifically for broad-band face images with aligned SFs (as compared to phase scrambled stimuli). The MEG results (Petras et al., 2021) further showed a reduced gamma range response in the early visual cortex, suggesting a reduction in HSF processing load for intact face stimuli. This indicates that the presence of informative LSF information facilitated the integration of HSFs. Building upon this research, we previously conducted a neuroimaging study investigating the impact of disrupting coarse-to-fine integrative processes using backward masking (Schuurmans et al., 2023). We utilised face stimuli known to elicit robust responses throughout the ventral visual pathway and employed backward masking to selectively disrupt LSF or HSF content of the stimuli at different time intervals. The masking process allowed for the disentanglement of SF contributions to neural responses while preserving the aligned SF scales present in natural input. In V1, ventral (i.e. FFA) dorsal, orbitofrontal brain regions, we found that disruption of LSFs through masking initially had a strong disruptive effect that decreased over time, while the opposite trend was observed for HSF masking. This pattern of activity supported the coarse-to-fine theories proposed earlier. However, it is important to note that while this study examined coarse-to-fine processing during face processing, it did not specifically investigate its contribution to perceptual behaviour.

The current study uses SF backward masking in order to investigate whether the disruption of coarse-to-fine integration has an impact on face identity recognition performance. We hypothesised that behavioural performance should mirror the processing profile observed in the FFA (Schuurmans et al., 2023). We expect slow and sustained processing of HSFs to rely more on recurrent processing, and makes them more susceptible to backward masking, leading to a steep increase in performance as a function of masking times. On the other hand, the rapid and transient processing of LSFs should be less affected by backward masking, resulting in a more gradual improvement in performance across masking durations. Full-spectrum face images of familiar celebrities were presented for 50, 75, 100, 150, or 300ms, directly followed by Fourier phase scrambled masks containing either LSFs ($< 11.4\text{cpf}$) or HSFs ($> 11.4\text{cpf}$). These cut-offs ensured that the SF range critical for face identity recognition was equally distributed across both LSF and HSF masking conditions to avoid any bias in the results towards either range. To quantify performance, the faces were embedded in Fourier phase scrambled noise and the signal-to-noise ratio (SNR) was manipulated in a staircase (i.e. lower SNR with a correct response, and vice versa). Through a linear mixed effect modelling (LMM) analysis, we analysed the SNR thresholds required for 75% accuracy on the identity matching task to examine the influence of masking LSFs and HSFs on face identity recognition performance. As expected, in both LSF and HSF masking conditions, the SNR threshold decreased as stimulus onset asynchrony (SOA) increased. This trend aligns with the prediction that extended processing time enhances the ability to recognise identity. However, we did not observe a steeper decline in performance for HSF masking as compared to LSF masking. These inconclusive results may be attributed to various factors, such as the use of familiar faces in our study, potentially relying stronger on and so increasing the influence of LSF information (e.g., Harmon, 1973; Sinha, 2002). These factors, along with the implications of the null-results, are explored further in the discussion, but a more comprehensive understanding requires further investigation.

Methods

Subjects

For now, the study included 19 healthy, naïve subjects (age: 23 ± 3.37 , 5 males) to participate in two experimental sessions. Prior to the testing, visual acuity of the subjects was evaluated using the Freiburg Visual Acuity and Contrast Test (FrACT), which demonstrated that all subjects possessed normal or corrected-to-normal vision (LogMar mean = -0.09 ± 0.12 , decVA = 1.33 ± 0.32). All subjects, except two, reported right-handedness according to the Edinburgh handedness Questionnaire (68.05 ± 51.69 ; Oldfield, 1971). Finally, all subjects performed adequately on the Benton task for facial recognition (mean score = 44.52 ± 2.93 ; Benton, 1980). Subjects were remunerated either with credits or a monetary compensation of 10 per hour. All subjects provided written informed consent, and the Ethical Committee of the Université Catholique de Louvain granted approval for the study (CEHF 2016/13SEP/393).

Stimuli

The principal experimental stimuli were generated using grayscale photographic images of the faces of 18 renowned male celebrities, with each celebrity represented by four photographs. An elliptical aperture was superimposed on each image to exclude the background. Each face was then resized to 250 x 200 pixels to subtend a visual angle of $6.3^\circ \times 5.1^\circ$, and centred on a uniform grey background of 550 x 550 pixels. To mitigate potential loss of extreme intensity values in the image, we rescaled the image luminance such that pixel values ranged from 0.1 to 0.9. Subsequently, we normalised the image luminance with a mean of zero and an RMS contrast of 1.

To create scrambled noise and mask stimuli, we employed phase scrambling, which preserves the amplitude spectrum of the original image while disrupting shape and edge information, thereby reducing image recognisability. Specifically, we utilised a custom MatLab function to randomise the phase structure of images in the Fourier domain. However, this technique creates an unintended increase in LSF energy in the presence of a uniform background, affecting the whole image (face and background). Therefore, we used an iterative scrambling procedure proposed by Petras et al. (2019) involving 500 iterations to counteract this effect. This method involves repeating the phase scrambling process and overlaying the original face pixels onto the phase-scrambled image over 500 iterations. The resulting image has a similar power spectrum in both the face and background regions (Jacobs et al., 2020; Petras et al., 2019, 2021). We utilised iterative phase scrambling to generate the scrambled copies of face stimuli ($n=72$) and scrambled noise stimuli ($n=8$).

The 72 phase-scrambled copies of face stimuli were used to create mask stimuli that were filtered to exclusively contain either LSF or HSF components. This enabled the targeted disruption of LSF or HSF processing in the full-spectrum face. To ensure that the mask images contained only LSF or HSF information,

we utilised 2nd order Butterworth filters with cut-offs that preserved four octaves for both SFs and maintained no gap between the two SF cut-offs (LSF: 0.71 - 11.36cpf; HSF: 11.36 - 500cpf).

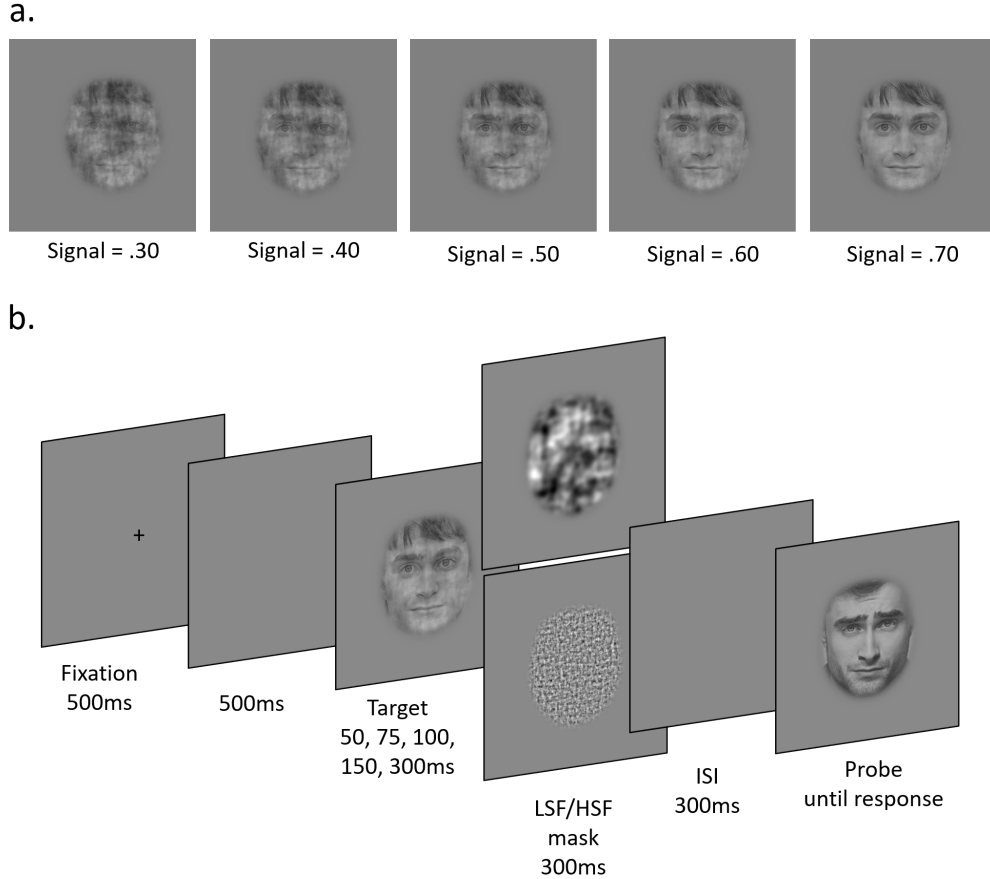


FIGURE 3.2: Face blending example and trial structure. a) An example of a face blended seamlessly into the background at various signal strengths. b) Illustration of a typical trial sequence - beginning with a fixation cross, followed by a display of the target face at varying SOAs, and then a phase scrambled mask containing only LSFs or HSFs. After an Inter-Stimulus Interval (ISI), the second face (probe) appears and remains until a response is registered.

To avoid edge artefacts, we multiplied face-shaped stimuli and masks with an aperture that transitioned gradually (from 0 to 1). A complementary aperture was multiplied with the uniform grey background to avoid contrast increases in the transition region. To provide a smooth transition between trial events, all stimuli in a block were alpha-blended with a scrambled noise stimulus (e.g. see Ales et al., 2012; Schuurmans et al., 2023). This alpha blending technique created a weighted sum of the stimulus and scrambled background contrast, while maintaining mean luminance. The weight of the stimulus contrast varied depending on the subject's response (Figure 3.2a). The final generated stimuli consisted

of a face embedded in a uniform grey background and blended with a phase-scrambled noise stimulus with a mean luminance and RMS contrast of 0.3 ± 0.2 . The masks were not blended with the noise stimuli and were given a higher RMS contrast of 0.4 to ensure strong masking of the SF range they contained.

Procedure

Subjects were seated in a light-tight chamber with their head position stabilised by a chinrest, at a distance of 57 cm from a ViewPixx monitor (1920 x 1200 pixels, 120 Hz, 27 cd/m² maximum luminance). The monitor's luminance levels were calibrated with a ColorCAL MKII colourimeter (Cambridge Research Systems, Kent, UK) to ensure linearity. The experimental paradigm was implemented using PsychoPy (version 2020.2.4; Peirce et al., 2019) and Python 3.8.5.

In the first session, a pre-screening task was administered to determine the eight most familiar celebrities for each subject. Specifically, subjects were shown a photograph of a celebrity, followed by five names. One of these names corresponded to the celebrity in the photograph, while the other four were randomly selected celebrity names. Subjects were instructed to accurately and quickly identify the celebrity name that matched the photograph. The eight celebrities that were most accurately identified were selected for use in the main experiment.

The experimental paradigm involved a delayed identity matching task with backward masking. In the first practice round, subjects were presented with 10 trials in which they were instructed to fixate on the central fixation cross. After a 500ms delay, a full-spectrum face stimulus was presented for an SOA of either 50, 75, 100, 150, or 300ms. The face stimulus was then masked with phase-scrambled LSF or HSF stimuli for 300ms, followed by the presentation of a second face stimulus (different picture, but same or different identity) (Figure 3.2b). The second face stimulus remained on the screen until the subject provided a "Same" or "Different" response via a button press. Subjects were instructed to respond as accurately and quickly as possible. The second practice round, comprising 10 trials, followed a similar protocol to the first practice round, with the exception of a manipulation of face visibility. Specifically, we employed a random SNR approach, blending the face stimulus with the background image at different SNR levels. Subjects were required to achieve a minimum accuracy threshold of 75% in both practice rounds to qualify for participation in the main experiment. This criterion was implemented to ensure that subjects comprehended the instructions and maintained attentiveness throughout the experimental procedure. The main experimental phase comprised sixteen blocks, each consisting of 5 smaller blocks containing 20 trials per condition. The design of the main experimental phase shared similarities with the practice rounds, but incorporated a staircase procedure to dynamically adjust the SNR of the face stimuli. Initially, the SNR level was fixed at 75, and depending on the subject's response, it was adapted using a standard adaptive procedure with a one-up two-down staircase (Leek, 2001) with step sizes of 0.5dB. This allowed us to determine the SNR level at which subjects could achieve a performance level of 80%. In case of pronounced disruption

by backward masking, the SNR level was expected to be higher at the end of the experiment. Conversely, if the masking had less disruptive effects, the final SNR level would be lower, indicating better performance. Two separate staircases were implemented per condition, resulting in a total of 160 trials per condition. Each staircase consisted of 80 trials, half of which were "same" and the other half "different". To improve the subsequent fitting of psychometric functions, every 10th trial was administered as a "jitter" trial (i.e., the SNR level was decreased or increased by 10%).

Analyses

In our analysis procedure, we first applied a logarithmic transformation to the response times to ensure a normal distribution. We then proceeded to exclude trials where response times exceeded three standard deviations from the mean within each condition for individual subjects. This approach aimed to remove outlier trials that might distort our results. As a result, an average of 2.1% of trials ($SD = 1.1\%$) per condition was excluded for each subject.

Psychometric fitting

In the psychometric fitting stage, we analysed the binary response data from our experiment (Figure 3.3, left panel), comprised both correct and incorrect responses, to identify the SNR corresponding to a 75% accuracy rate. Data from each subject was evaluated separately, considering different combinations of SOA and SF masking conditions. The psychometric fitting was carried out using a generalised linear model (GLM) with a specific logit link function tailored for our multi-alternative forced-choice (mafc) scenario. The logistic model took the form:

$$\frac{CorrectResponses}{(CorrectResponses + IncorrectResponses)} = \text{logit}^{-1}(\text{Intercept} + SNR) + 0.5$$

Where 'logit⁻¹' denotes the logistic function and 'Intercept' is the estimated constant term from the model. The guess rate of 0.5 is added to account for the probability of subjects guessing the correct answer. The graphical representation of this model is displayed in the right panel of Figure 3.3, where (binned) data points correspond to the proportion of correct responses at various SNR levels, and the overlaid curve indicates the model's fit to the data. The vertical dotted line represents the SNR threshold necessary for 75% accuracy.

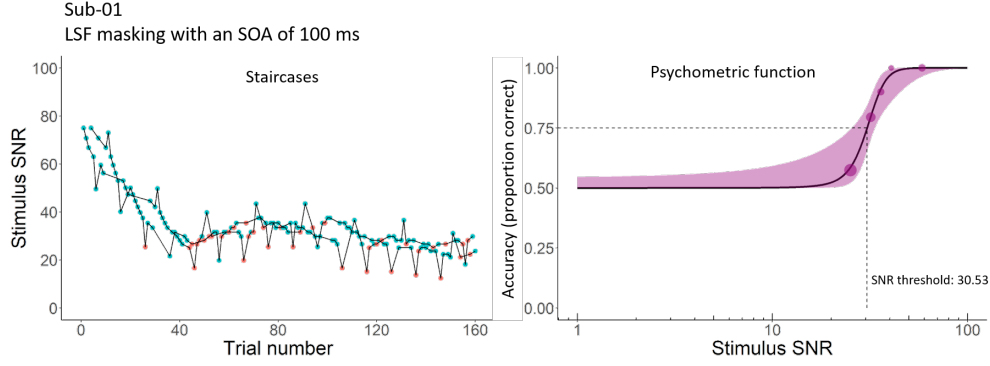


FIGURE 3.3: Staircase and psychometric function of a representative condition and subject. The left panel displays the evolution of two distinct staircases throughout the experiment, with blue markers designating correct responses and red markers indicating incorrect responses. The right panel exhibits the corresponding psychometric function. Data points, grouped into five bins, signify the data, with their size proportional to the frequency of presentation at each corresponding stimulus SNR. The overlaid curve denotes the model fit to this data, encapsulated by shaded bands representing 95% confidence intervals. The vertical dotted line indicates the SNR threshold required to attain 75% accuracy, serving as a measure of perceptual efficacy under the illustrated experimental condition for this representative subject.

Linear Mixed effects Modeling

A LMM was employed in R, and a Type III ANOVA using Satterthwaite’s approach was conducted to examine potential variations in threshold levels across conditions. In order to establish the most parsimonious model, we employed a stepwise variable selection procedure with backward elimination, following the methodology outlined by (Heinze et al., 2018). Initially, a full-model model was formulated, including interaction effects - this model was employed for data visualisation. Subsequently, we implemented a systematic elimination of non-significant effects (two-way interaction/main effect). This step-by-step simplification aimed to establish a minimal yet efficient model that optimises the analytical power. The derived model’s adequacy was affirmed by verifying if it produced the smallest Bayesian Information Criterion (BIC), as lower BIC values indicate a better balance between fit and simplicity. The resulting model for data analysis included only the main effect of SOA (BIC = 1212.73):

$$threshold \sim 1 + soa + (1|subject)$$

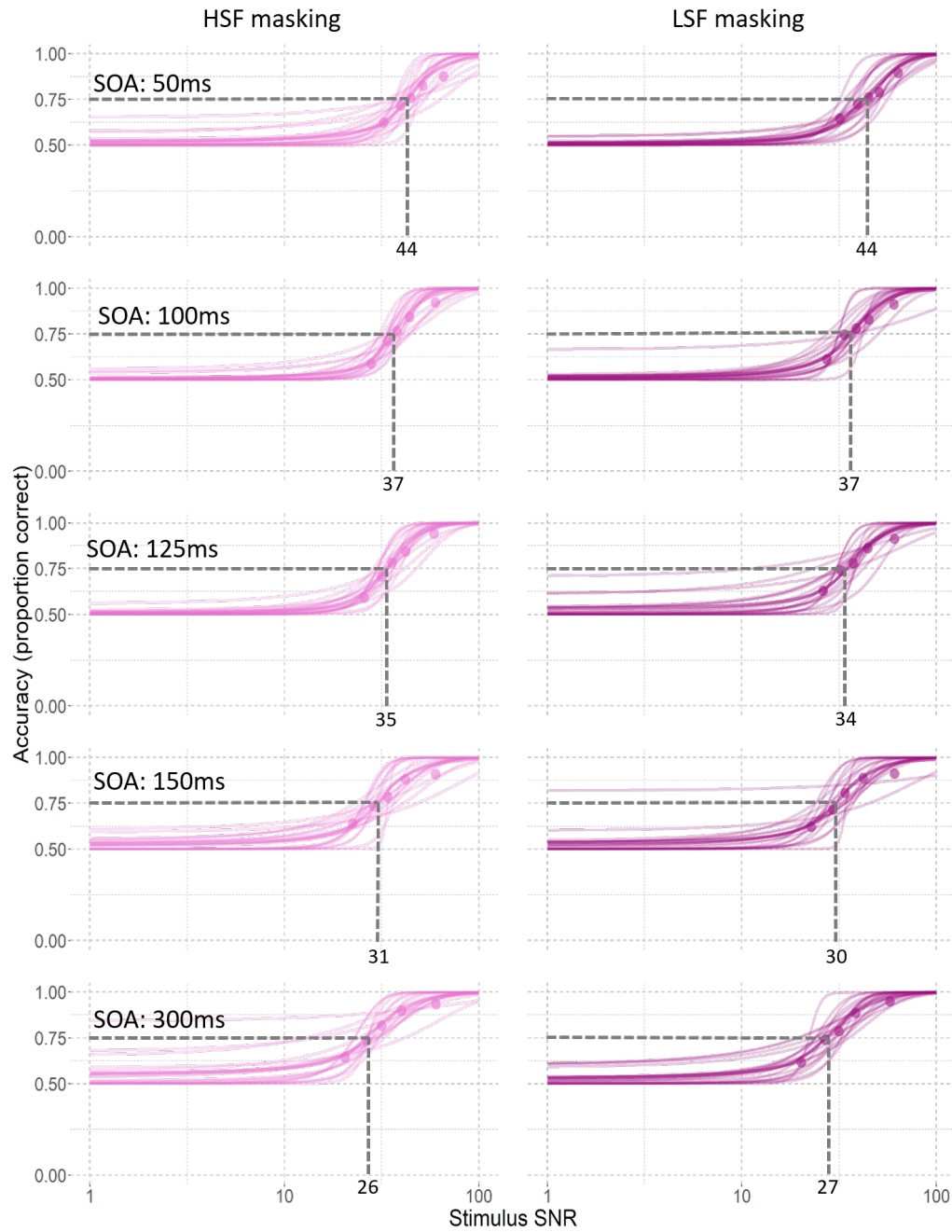


FIGURE 3.4: Logistic psychometric fitting. The graphs showcase the individual psychometric functions denoted by thin lines. Marker points (coloured dots) represent the averaged correct response rate across all subjects. The thick coloured line in the plot depicts the group's average psychometric function, providing an estimate of the group-level performance across the range of stimulus parameters. The dotted line illustrates the performance level at 75% correct responses and corresponding average SNR level.

Results

Psychometric fitting

The psychometric functions fitted to individual subjects' responses are represented in Figure 3.4. For comparison, group averages of correct responses, sorted by stimulus parameters, along with the corresponding psychometric function are depicted by thick dots and lines. The psychometric fitting process enabled us to pinpoint threshold values necessary for 75% correct responses in the different conditions, giving us an index of perceptual efficiency under these experimental manipulations.

Linear Mixed effects Modelling results

Based on the nature of coarse-to-fine processing of visual information, we hypothesised that both SOA and SF masking would play crucial roles in determining the identity recognition thresholds. Specifically, we expected that due to the slow, sustained processing of HSFs, HSF masking would lead to a pronounced increase in face recognition performance (i.e., decreasing SNR thresholds) as the SOA increases. In contrast, we expected the quick, transient processing of LSFs to be less influenced by backward masking, resulting in a less steep performance slope as a function of SOA.

After the systematic removal of non-significant higher-order terms of our LMM, the Type III ANOVA revealed a significant main effect of SOA ($F = 215.45$, $p < .001$). This main effect underlines our expectation of a negative slope in performance as a function of SOA, thus supporting our hypothesis that longer processing time enhances face recognition performance. Despite this, we found no significant difference between LSF and HSF masking conditions, contrary to our initial hypothesis. Individual differences accounted for a proportion of the total variance (5.5%), indicating that inter-subject variability could play a role in shaping the outcomes. The observed variance underscores the importance of considering individual differences, perhaps in the SF processing or in the face identification recognition strategies employed during the experiment. The average SNR thresholds and LMM fits, derived from a full model that includes all interaction terms for a comprehensive overview, are represented in Figure 3.5 and Table 3.1.

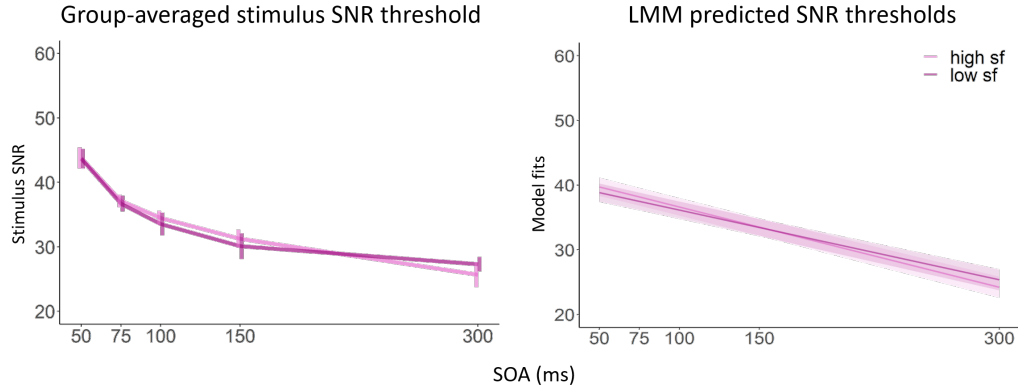


FIGURE 3.5: Results for stimulus SNR thresholds. The left panel illustrates the group-average stimulus SNR thresholds, encompassed within a 95% within-subjects confidence interval (Loftus & Masson, 1994). The right graph presents the predictive modelling outcomes derived from the LMM. Although there was no significant two-way interaction effect, these visualisations (of the full-model including a two-way interaction) are included for illustrative purposes, providing insights into the structure of the data and potential trends. The final model identified only a main effect attributed to SOA.

Several factors might have contributed to these null findings, including the close similarity of our experimental conditions or the familiarity of the presented faces that possibly heightened the role of LSF information (see Discussion).

TABLE 3.1: Model predictions (estimated marginal means - *emmeans*) of collapsed LSF and HSF masked conditions, for different SOAs.

SOA	<i>emmean</i>	<i>SE</i>	<i>df</i>
50	39.3	1.31	20.6
75	37.8	1.29	19.3
100	36.4	1.27	18.4
150	33.5	1.27	18.1
300	24.8	1.42	28.6

Discussion

The coarse-to-fine processing framework posits that LSF cues are processed before, and guide the slower processing of HSF information (Bar, 2004; Bullier, 2001; Hegdé, 2008; Kreiman & Serre, 2020; Watt, 1987). Indeed, our previous neuroimaging study (Schuurmans et al., 2023) showed that in early visual areas, ventral, dorsal, and orbitofrontal in the brain, early disruption of LSFs in face stimuli evokes high responses, decreasing over time, while the opposite occurred with HSF masking, supporting the coarse-to-fine theories. In light of this, our study was aimed at exploring the behavioural implications of coarse-to-fine processing, specifically focusing on how disruptions in this mechanism may affect face identity recognition. Drawing upon the coarse-to-fine literature, our hypothesis posited that LSF processing would be rapid and display reduced susceptibility to backward masking, whereas HSF processing would be slower and show heightened sensitivity to masking effects. To investigate this, we used images of familiar celebrities, presented in full-spectrum, but with either their LSFs or HSFs selectively disrupted through backward masking. The masking was deployed at increasing SOAs. While the results showed improved identity recognition performance as a function of SOAs, we did not observe any significant difference between masking different SF ranges. It is unclear whether the deviation from our expected findings is due to the particular task design not being sensitive enough to capture coarse-to-fine identity recognition, or whether they should be taken to challenge the conventional coarse-to-fine framework.

Our experiment leveraged full-spectrum facial stimuli consistent across masking conditions. The only variable was the SF content of the masks, resulting in a subtle difference between conditions. This approach, while elegant, provided a conservative test for the contribution of coarse-to-fine processing to face identity recognition. We utilised a method similar to the approach taken in our previous neuroimaging study (Schuurmans et al., 2023); masking specific SF ranges of a target stimulus to constrict their processing after various durations. The fMRI study yielded robust differences in neural responses to various SF masking conditions, resonating with the principles of coarse-to-fine processing. However, these modulations were absent in the current behavioural study, raising questions about whether the subtlety of the differences between conditions inhibited the detection of masking effects in an identity recognition task. It leads us to contemplate whether adjusting the strength and properties of the mask might yield more pronounced results. Our masking choices drew inspiration from a study by Bacon-Macé et al. (2005), where they conducted a behavioural categorisation experiment, presenting images from various categories followed by noise masks. Subjects were tasked with identifying whether the image represented an animal. As in the current study, their masks were of varied spatial resolutions, and they intentionally used masks with maximal contrast to ensure strong masking interference (Albrecht et al., 2002; Foxe & Simpson, 2002; Nowak et al., 1995), thereby robustly masking the original stimulus. In our experiment, we opted for mask stimuli with a slightly higher contrast than the preceding face stimuli but

strived for spectral similarity. Despite our efforts, the contrast increase might not have been sufficient. Bacon-Macé et al. (2005) also experimented with a second masking condition, a "dynamic mask", consisting of sequential noise masks with very brief durations. This was employed due to its apparent efficiency, evidenced by performances at chance level with the shortest SOA. Even though a dynamic mask is more disruptive than a static mask, this method also introduces a potential issue: the uncertainty of pinpointing the exact moment at which the masking becomes fully effective, possibly leading to substantial ambiguity regarding the disruptive latency. This raised a consideration for our study regarding the use of a dynamic mask to achieve more effective masking. However, in a pilot version of our research, we found no significant alteration in the dynamics of the results, leading us to favour a static mask approach. Another avenue worth exploring might be the employment of various types of masks. In an experiment concerning identity matching of human faces, Bachmann et al. (2005) investigated backward masking effects using different types of masks: face masks that differ in identity from the initial face and Gaussian noise masks. The results underscored that face masks induced more effective backward masking than noise masks. The distribution of contrast within these mask types, such as the strong contrast around the eye region in face masks compared to the random spread in noise masks, may explain this discrepancy. Although our study observed an effect of masking (with stronger effects for short SOAs decreasing over time), the masking might need to be more pronounced especially around the eye region to detect differences between SF masking conditions. Therefore, our future work might benefit from adopting the usage of face masks to enhance the masking of initial face stimuli.

The influence of masking on our results could potentially be attributed not only to its strength but also to the SF range of the masks. Research has consistently associated identity recognition of faces with SFs in the range of 8-16cpf, particularly focusing on a mean frequency of 11cpf (Costen et al., 1996; Gold et al., 1999; Näsänen, 1999; Parker & Costen, 1999). Guided by this research, we opted for an SF cut-off of 11.364cpf for the noise mask, with the intention of including some of the SFs that are crucial for face recognition in both mask types. Despite our careful selection of SF range, the lack of observable differences between SF masking conditions, raises concerns about the efficacy of the chosen SF cut-off. A point of divergence arises when comparing the current study to our previous neuroimaging research (Schuurmans et al., 2023), where we utilised a lower SF cut-off (5.7cpf). This divergence implies that the current study's masks contain a greater presence of higher SFs. While the SF cut-off in terms of the cycles per degree (cpd) of visual angle maintained a closer proximity between the neuroimaging study (1.8cpd) and the current study (2.2cpd). Several studies have underscored the stability of identity recognition across a broad spectrum of stimulus sizes (Andrews & Ewbank, 2004; Y. Lee et al., 2006; Oruç & Barton, 2010; L. Zhao & Chubb, 2001), which implies that cpf could be more relevant than cpd of visual angle. We had initially anticipated a pronounced masking effect with HSF masks. However, when both masks are skewed towards HSFs

in terms of cpf, any potential differences between SF masking conditions could be obscured. Thus, the disparity in SF cut-off between our two studies could potentially account for the differing outcomes observed in this behavioural study.

The divergence in methodology between our current study and the previous neuroimaging research extends beyond merely the chosen SF cut-off; it also encompasses the types of faces used. In the neuroimaging study, unfamiliar faces were employed, whereas in the current investigation, we deliberately utilised familiar (celebrity) faces to engage authentic face recognition mechanisms that can differentiate between varying faces (e.g., Andrews and Ewbank, 2004; Kramer et al., 2018; Oruç and Barton, 2010). Unfamiliar face recognition tends to rely on image-matching (Hancock et al., 2000; Megreya & Burton, 2006), where the comparison between images of the same unfamiliar person does not typically engage face-related processing, but rather on pixel-level scrutiny, applicable to any visual pattern (Burton, 2013). To focus on genuine face-related processing, we opted for familiar faces in our experimental design. The recognition of familiar faces, often possible even at a single glance (Barragan-Jason et al., 2015; Caharel et al., 2014; Hsiao & Cottrell, 2008; Visconti di Oleggio Castello & Gobbini, 2015; Yan & Rossion, 2020), suggests a reliance on LSFs, which capture the global face structure (Goffaux & Rossion, 2006; C. H. Liu et al., 2000). Despite a broad consensus about the importance of mid SF ranges in face recognition (Costen et al., 1994; Oruç & Barton, 2010; Yan et al., 2022), familiar faces have been recognised even at lower spatial resolution (Ginsburg, 1980; Harmon, 1973; Harmon & Julesz, 1973; Rubin & Siegel, 1984; Sinha, 2002; Yip & Sinha, 2002). For example, a recent study by Yan et al. (2022) found that the neural response for familiar face recognition emerges at about 6.3-8.7cpf, indicating that such recognition can occur at coarser visual inputs than traditionally considered. This observation aligns with work of Dobs et al. (2019), where they found an enhancement of identity information for familiar faces compared to unfamiliar faces during the early stages of processing. The implication is significant for our understanding of SF processing: if early processing stages are predominantly engaged with LSFs, and if familiar faces elicit enhanced representations earlier than unfamiliar faces, it suggests a stronger reliance on LSFs for the processing of familiar faces. This early enhancement, driven by LSFs, enables the brain to rapidly leverage the global structure of familiar faces, facilitating quicker and more efficient recognition. The prioritisation of LSFs in familiar face processing might explain why familiar faces are recognised even at lower spatial resolutions. Thus, in the context of our study, the chosen SF cut-off of 11cpf might not have sufficiently differentiated between the LSFs and HSFs critical for the recognition of familiar faces. As familiar faces could be recognisable using information below this threshold, the overlap in relevant identity cues across both LSF and HSF ranges in our SF masks could have diminished the perceptual distinction between the masking conditions, potentially obscuring the distinct contributions of low and high SFs to processing. This might have led to the reduced differentiation between LSF and HSF masking that we observed. The reliance on higher SFs for processing unfamiliar faces, given that they provide the finer details critical for distinguishing

between unfamiliar individuals (Goffaux & Rossion, 2006; C. H. Liu et al., 2000; Visconti di Oleggio Castello & Gobbini, 2015), implies that conducting an experiment utilising unfamiliar faces might result in a different pattern of findings. It could emphasise the sustained role that HSFs might play in the coarse-to-fine processing framework. Thus, the utilisation of familiar faces in our study might have altered the results, leading to an absence of observable differences between SF masking conditions.

In the current study, we strategically employed familiar faces to steer subjects away from image-level matching. However, the choice of a face matching task may have inadvertently complicated the results, considering that the task itself can significantly sway the range of SFs critical to the visual system (Schyns, 1998). While face identity recognition and identity matching tasks are related, they arguably represent distinct cognitive processes that might invoke different neural mechanisms (Biederman & Cooper, 1992; Sergent & Poncet, 1990; Young et al., 1993) and engage different brain regions (Benton, 1980; Leveroni et al., 2000; Rossion et al., 2001; Warrington & James, 1967; Young et al., 1993). Face identity recognition denotes an automatic, almost effortless process for most people, often relying on mid-SF information essential for distinguishing individual facial features (Costen et al., 1996; Gold et al., 1999; Näsänen, 1999; Parker & Costen, 1999). Contrarily, identity matching, a more deliberate process, might require finer HSF details, particularly in distinguishing between closely resembling faces (Bachmann, 1991; Bindemann et al., 2013; Costen et al., 1994, 1996). The use of a matching task in our study, which might accentuate the role of HSF information, stands in contrast with the use of familiar face stimuli that are typically associated with LSF processing. This discrepancy between task and stimulus characteristics could have introduced a misalignment in the study's methodology, affecting the interpretability of our findings. Specifically, the matching task may have promoted a more detailed HSF-focused approach, while the familiar faces might have engaged lower SF processing. This discordance could have masked the specific roles of low and high SFs in face recognition, reducing the anticipated differences between the two SF masking conditions. A different task, more aligned with identity recognition, might have yielded alternative insights into the coarse-to-fine dynamics in face identity recognition.

The absence of an effect of SF masking in our study may appear to challenge the coarse-to-fine framework. While numerous neuroimaging studies have lent support to the dynamics of coarse-to-fine processing (e.g. Bar, 2004; Bullier, 2001; Goffaux et al., 2011; Kauffmann, Bourgin, et al., 2015; Petras et al., 2021; Schuurmans et al., 2023), it is essential to recognise that these findings may not necessarily translate to observable behavioural responses. The relationship between neural activity and behavioural output is multifaceted, and factors such as experimental design, stimulus properties, and individual differences can all influence the manifestation of an effect at the behavioural level. However, the scientific community must approach this absence of evidence with caution. It is a fundamental principle in scientific inquiry that the lack of a significant test result does not equate to an absence of the effect itself (Greenland et al., 2016).

Null results can stem from various sources, including methodological limitations, insufficient power, or the subtlety of the effect in question. It is also worth noting that individual variations in processing or sensitivity to masking could further obfuscate the results. The observed inter-subject variance, accounting for 5.5% of the variability in our LMM analysis, underscores the importance of individual differences in SF processing or face recognition strategies. The variance indicates that individual processing styles or sensitivity to SF masking could play a role in shaping the outcomes. The ANOVA, not accounting for these individual differences, could overlook subtle yet crucial variations in processing. This potential oversight highlights the need for future research to consider these individual variances, which might reveal more nuanced insights into the role of SFs in face recognition.

In conclusion, our investigation into the coarse-to-fine processing for human face identity recognition through backward masking has opened a dialogue on the sensitivity of the current design to test dynamics of face identity recognition. The lack of results supporting the coarse-to-fine hypothesis could be attributed to factors such as our chosen SF cut-off, the use of familiar faces, and the nature of our masking approach. Future research might benefit from alternative masking techniques, varying task demands, and more fine-grained manipulation of SFs to address the contribution of coarse-to-fine processing to face identity recognition behaviour. The study serves as a reminder that even tested theories can contain elements that are yet to be fully understood, inspiring continued inquiry into the complex world of human visual cognition.

Chapter 4

Enhanced recurrent processing under conditions of visual input degradation

Jolien P. Schuurmans¹, Valérie Goffaux^{1,2}

¹Psychological Sciences Research Institute, Université Catholique de Louvain, Louvain-la-Neuve, Belgium; ²Institute of Neuroscience, Université Catholique de Louvain, Louvain-la-Neuve, Belgium;

Abstract

Visual perception relies on recurrent communication within and across different levels of the cortical hierarchy. This iterative process allows for the integration of information between early visual areas (V1/V2) and higher-level regions, enabling the extraction of relevant features and the suppression of irrelevant ones in V1/V2. Previous studies have demonstrated with computer models, monkey physiology or behaviourally in humans that recurrent processes are stronger when a stimulus is noisy or degraded. This interaction between stimulus degradation and recurrent processing in the human visual system remains unclear. This study aimed to investigate the involvement of the V1/V2 and other visual regions (i.e. fusiform face area, FFA and occipital face area, OFA) in recurrent processing and their connectivity under conditions of stimulus degradation. We used an fMRI backward masking paradigm to disrupt recurrent processing of human face stimuli at various stimulus onset asynchrony (SOA) timings. Stimuli were presented under varying levels of signal-to-noise ratios (SNRs). Based on the hypothesis that recurrent processing increases as a function of stimulus noise, it was predicted that high SNR stimuli would be rapidly processed, resulting in reduced susceptibility to backward masking over time. Conversely, low SNR stimuli were expected to require more recurrent processing and to show an increased vulnerability to backward masking. We indeed found an interaction between SOA and SNR in V1/V2, reflecting the modulation of recurrent processing, as indexed by the influence of backward masking over time, based on stimulus strength. These findings yield new insights on the mechanisms underlying visual perception and recurrent processing in the human visual system. By elucidating the dynamics of recurrent processing under conditions of stimulus degradation, this study contributes to our knowledge of how the visual system adapts to challenging perceptual environments.

Introduction

As one of the primary senses, vision plays a critical role in how we interact with and interpret the world around us. Visual processing relies on inter- and intra-level communication in the cortical visual hierarchy (Gilbert & Li, 2013; Hupé et al., 1998; Kravitz et al., 2013; Kreiman & Serre, 2020; Lamme et al., 1998; Petro et al., 2014). Besides the feed-forward transmission of visual signals (i.e. bottom-up), feedback mechanisms facilitate integration of information from high- to low-level regions (Lamme & Roelfsema, 2000), and horizontal connections foster local information assimilation (i.e. Angelucci and Bressloff, 2006). The process of feedback modulation is an essential component of recurrent processing, creating a loop of information exchange in the brain (Pennartz et al., 2019).

Upon receiving visual input, the signals travel through various regions, with the primary visual cortex (V1) serving as an initial processing stage. V1 and V2 are known to be sensitive to simple stimulus properties such as edges at specific orientations, and spatial frequencies (K. K. De Valois et al., 1979; Felleman & Van Essen, 1991; Hubel & Wiesel, 1962). From V1/V2, feed-forward projections transport visual signals to higher-order visual regions in the processing hierarchy in the ventral and dorsal pathways. The feed-forward transmission of visual signals travel from the occipital regions via the parietal lobe (within the dorsal, 'where?', pathway) to the orbitofrontal cortex, which is thought to activate plausible interpretations of the stimulus (Bar et al., 2006; Bar, 2004; Kauffmann, Bourgin, et al., 2015; T. S. Lee, 2015; Mumford, 1992). Concurrently, information is relayed to the ventral (or temporal, 'what?') pathway, which is sensitive to complex visual categories such as objects, faces, and scenes (i.e. fusiform face area, FFA; parahippocampal place area, PPA). The signals in these selective visual regions are proposed to prompt the activation of an internal template representation of the stimulus, modulated not only by feed-forward input, but also recurrent feedback signals from areas such as the orbitofrontal and parietal regions (Ayzenberg & Behrmann, 2022; Ayzenberg et al., 2023; Bar et al., 2006; Tomita et al., 1999; Van Dromme et al., 2016). This continuous feedback, occurring even during the feed-forward transition, is directed from higher-level regions in the temporal, parietal, and frontal regions towards V1/V2 (Ayzenberg & Behrmann, 2022; Felleman & Van Essen, 1991; Kveraga et al., 2007; Zachariou et al., 2017), enhancing relevant aspects of the visual stimulus while suppressing irrelevant ones.

While feed-forward processes might suffice for processing simple, isolated stimuli, recurrent computations become critical when faced with more complex or demanding conditions with degraded or ambiguous input (Groen et al., 2018; Kar et al., 2019; Kreiman & Serre, 2020; Rajaei et al., 2019; Seijdel et al., 2020; Spoerer et al., 2018; Tang et al., 2018). Such conditions introduce variability and uncertainty, thereby challenging the ability to accurately interpret these noisy signals. When stimuli present a low signal-to-noise ratio (SNR), the brain presumably leans heavily on recurrent processes to mitigate this variability and enhance perceptual robustness (O'Reilly et al., 2012; O'Reilly et al., 2013; Rajaei

et al., 2019; Wyatte et al., 2012). Studies in non-human primates have shown that degraded stimuli (e.g., in case of occlusion) necessitate additional recurrent processes to extract relevant information and make accurate perceptual judgments (Kar & DiCarlo, 2021; Kar et al., 2019; Kovács et al., 1995; Nielsen et al., 2006). Studies have also shown that simple stimuli with low contrast (such as gratings or noise patches) make neurons in V1 more susceptible to modulation through lateral or top-down recurrent processes (Krause & Pack, 2014). In a behavioural study, Wyatte et al. (2012) explored the function of recurrent processing in visual object recognition. They degraded stimuli through occlusion and contrast degradation and manipulated recurrent processing mechanisms using backward masking (Enns & Di Lollo, 2000; Lamme & Roelfsema, 2000). Their findings showed a significant impairment in categorisation performance when a mask followed a heavily degraded stimulus, which disrupted recurrent processing dynamics. However, performance remained stable when presented with a clear stimulus. These observations imply that higher-order visual signals are needed to guide the visual processing and to categorise the distorted visual input (O'Reilly et al., 2012; Rajaei et al., 2019; Wyatte et al., 2012). Computational models lend further support to these observations (O'Reilly et al., 2013), showing that when stimuli are degraded, the bottom-up signals are weak and noisy, leading to an ambiguous pattern of activation across the highest levels of the hierarchy. However, through recurrent propagation of signals between adjacent layers, these signals mutually reinforce each other despite the degraded bottom-up stimulus, ultimately leading to a robust and stable pattern of activation at the higher processing levels. Despite converging evidence from human, non-human primate, and computational studies, the specific role of recurrent processing in managing complex, degraded stimuli within the human visual system remains unclear. In particular, the contribution of primary visual regions such as V1 and V2, proposed to be central in the recurrent exchange of brain signals for visual perception (Bullier, 2001; T. S. Lee, 2015; Mumford, 1992), is elusive. Addressing this gap requires an examination of the relationship between stimulus degradation and recurrent processing in human V1/V2, focusing on naturalistic stimuli such as human faces, for which the visual system has developed robust template representations.

In the present study, we sought to explore the engagement of V1/V2 and higher-level visual regions in recurrent processing of faces under conditions of stimulus degradation. Given the findings from prior research, our hypothesis posited that the degree of recurrent processing would increase in response to the presence of noisier or more degraded complex visual stimuli in V1/V2. We expect higher-level visual regions to respond based on their specific selectivity, for instance, the FFA/OFA would respond strongly to face stimuli. However, the introduction of noise to face stimuli is likely to reduce this sensitivity, resulting in subdued responses (e.g., Liu-Shuang et al., 2022; Schuurmans et al., 2023). As the need for recurrent processes heightens, we anticipate increased correlation between V1/V2 and FFA/OFA. To put these hypotheses to the test, we opted for face photographs as stimuli. Faces elicit internalised template representations within higher visual face selective areas in the ventral regions (Schuurmans et

al., 2023). By embedding these face stimuli within varying levels of noise, we established different SNRs. High SNR stimuli featured clearer, less noisy face information, whereas low SNR stimuli were more degraded and presented more of a perceptual challenge. The strategic combination of potent stimuli and backward masking, a manipulation known to disrupt recurrent processing (Bridgeman, 1980; Enns & Di Lollo, 2000; Kovács et al., 1995; Lamme et al., 2002; Macknik & Livingstone, 1998; Rolls & Tovee, 1994; Rolls et al., 1999; Thompson & Schall, 1999; Wyatte et al., 2012), created an optimal platform to test our hypotheses and unravel the extent and dynamics of recurrent processing. In our study, subjects were presented with face stimuli immediately followed by a phase-scrambled mask. This backward masking approach, coupled with the application of fMRI, provided the tools to examine temporally sensitive queries with high spatial resolution. Disrupting recurrent processing at varying stimulus onset asynchronies (SOAs) - set at 50ms, 75ms, 100ms, and 150ms (timing adapted from Goffaux et al., 2011; Schuurmans et al., 2023) allowed us to elucidate the temporal dynamics of recurrent processing specifically within V1/V2 and higher visual face selective regions (i.e. FFA and OFA) under different levels of stimulus degradation.

We expect that in V1/V2 an increase in SOAs will allow more time for recurrent processes, leading to sparser representations and reduced neural responses. For high SNR stimuli, which are processed rapidly due to their clarity, we expect overall low activity and a modest decline in responses as a function of SOA, reflecting a reduced need for recurrent processing and a lower sensitivity to the effects of backward masking. In contrast, the decline is predicted to be more pronounced for low SNR stimuli. These stimuli, due to their complexity and noise, rely more heavily on prolonged recurrent processing for accurate interpretation. As such, backward masking is likely to have a substantially disruptive effect on them. We indeed found an interaction between SOA and SNR within V1/V2; a stronger modulation of SOA for low SNR compared to high SNR stimuli. The initial high response with low SNR masking was lower for masking as a later SOA. This effect was not as pronounced for high SNR masking. In face-selective regions, we recorded increasing responses as a function of SOA, with overall higher responses for stimuli with a higher SNR, underscoring the regions' established tendency to respond more robustly to intact faces. In an exploratory beta series correlation analysis, we examined interactions between various ROIs, including V1/V2, FFA, and OFA, under different SOA and SNR conditions. We found negative correlations between V1/V2 and FFA across all conditions, suggesting compensatory activation patterns where an increase in V1/V2 activity correlates with less activity in the FFA when the stimulus is ambiguous. This negative correlation appeared to decrease slightly as SOA increases, especially for low SNR stimuli. Conversely, between V1/V2 and OFA, we found an increase in positive correlations, suggesting strengthened connectivity. This effect was more pronounced at later SOAs, and was particularly evident in high SNR conditions. FFA and OFA displayed overall negative correlations, possibly indicating competition between the regions, with this effect more pronounced for high SNR stimuli and extended SOA (see discussion). This preliminary analysis

is set to be supplemented with a more refined Psychophysiological Interaction (PPI) analysis in due time. The insights obtained from these findings contribute to our understanding of the dynamic role of recurrent processing under varying levels of stimulus degradation, thus offering a deeper understanding of how the visual system adapts to varying perceptual challenges.

To conclude, the insights gained from this study not only validate the fundamental role of recurrent processing during the processing of complex stimuli in the visual system but also shed light on its dynamic and adaptive nature under different stimulus conditions/degradations. While our initial findings present a promising start, future research employing more nuanced techniques like PPI analysis promises to further our understanding of recurrent processes during visual perception in the human brain.

Methods

Subjects

Ten healthy subjects (age 28 ± 2.76 , 3 males), without any history of neurological disorders, underwent two scanning sessions. Visual acuity was assessed using the Freiburg Visual Acuity and Contrast Test (FrACT) before scanning. The results indicated that all subjects had normal or corrected-to-normal vision (LogMar mean = -0.01 ± 0.14 , decVA = 1.44 ± 0.36). All subjects self-reported as right-handed, with a mean Edinburgh handedness questionnaire score of 94.3 ± 15.12 (Oldfield, 1971). Additionally, all subjects demonstrated satisfactory performance on the Benton task for facial recognition (mean score = 47.7 ± 2.81 ; Benton, 1980). Prior to participation, all subjects provided written informed consent. Ethical approval for the study was granted by The Scientific and Ethical Review Board of the Faculty of Behavior & Movement Sciences, VU University Amsterdam (VCWE-2020-004R1), as well as by the Ethical Committee of the Université Catholique de Louvain (CEHF 2016/13SEP/393).

Stimuli

All subjects completed both the main experiment and a functional localiser to map the face-specialised regions in the ventral pathway at the individual level. During both the functional localiser and main experiment, stimuli were presented with PsychoPy3 2022.2.4 on a 32" BOLD screen from Cambridge Research Systems (1920 x 1080 pixels; 70.9 x 39.8 cm; refresh rate = 120 Hz) located at the end of the scanner bore. The subjects viewed the stimuli using a mirror affixed to the head-coil while maintaining a viewing distance of 210 cm. The gamma function of the monitor was accounted for by linearising the images' intensity values. The luminance and contrast of the pixel values had a maximum luminance of 325 cd/m² and a root mean square (RMS) contrast of 0.1 in 0-1-pixel value.

Main experiment

During the main experiment, subjects were sequentially presented with face stimuli followed by Fourier phase-scrambled masks. The face stimuli were embedded into a phase-scrambled background and blended with phase-scrambled noise to manipulate the level of stimulus degradation. The main experimental stimuli consisted of 20 greyscale photographs of unfamiliar white human faces (half male) captured with a neutral expression, using 367 x 267-pixel dimensions and subtending a visual angle of 3.7° x 2.7°. These images were sourced from young adult alumni students (aged 18-25 years) of the Université Catholique de Louvain (Belgium), who provided written consent for the use of their photographs (Laguerre et al., 2012). To prevent clipping of extreme intensity values in the images, their luminance was rescaled to ensure pixel values remained between 0.1 and 0.9. Subsequently, image luminance was normalised (mean of zero and RMS contrast of 1). The faces were then centred on a uniform grey background of 550 x 550 pixels (5.5° x 5.5° of visual angle). Each face appeared in three variations: high signal-to-noise ratio (SNR), low SNR, and phase-scrambled. We included the phase-scrambled condition to control for non-specific effects of visual stimulation. In order to create stimuli with different SNRs, we combined the face stimulus with a phase-scrambled image (as described in Ales et al., 2012; Schuurmans et al., 2023; and see Figure 4.1). All stimuli were masked by a phase-scrambled mask, and appeared on a circular phase-scrambled background (5.5° diameter). Therefore, we needed to produce several phase-scrambled copies of the face stimuli, to create the scrambled control condition (n=20), backward masks (n=20), and stimulus backgrounds (n=10). The phase-scrambled stimulus backgrounds were also used to create both high and low SNR stimuli (see below, and Figure 4.1).

We generated phase-scrambled images by randomising their phase structure in the Fourier domain using a custom MATLAB function. However, since manipulations in the Fourier domain affect the whole image, including the uniform background, the scrambled images might be contaminated by the uniform background. To reduce this confound, we employed an iterative phase scrambling

procedure (as in e.g. Jacobs et al., 2020; Petras et al., 2019, 2021), which involved scrambling the face image and then superimposing the original face back onto the scrambled image, and repeating this procedure 500 times. This iterative scrambling approach results in phase-scrambled images that have similar spectral properties as the face, thereby minimising the influence from the uniform background.

We employed alpha blending to combine the stimulus with the phase-scrambled image at various SNR levels (as described in Ales et al., 2012; Schuurmans et al., 2023; and see Figure 4.1). This technique produces a weighted sum of the contrast of the stimulus and the scrambled image while preserving the mean luminance. Specifically, an SNR of 35 indicates a weight of 0.35 for the stimulus and 0.65 for the noise. The low SNR value was determined for each subject using a quick pre-test, and in the main experiment, all subjects had a low SNR of either 30 or 35. The high SNR stimulus was consistently 30 units higher than the low SNR value. For instance, if a subject had a low SNR of 35, the high SNR stimulus would be 65. The moderate RMS contrast of the stimuli (i.e., their reduced weight in the final alpha-blended stimulus) produced challenging viewing conditions expected to recruit recurrent processing (Hupé et al., 1998; Kar & DiCarlo, 2021; Kar et al., 2019; Mohsenzadeh et al., 2018; Wyatte et al., 2014).

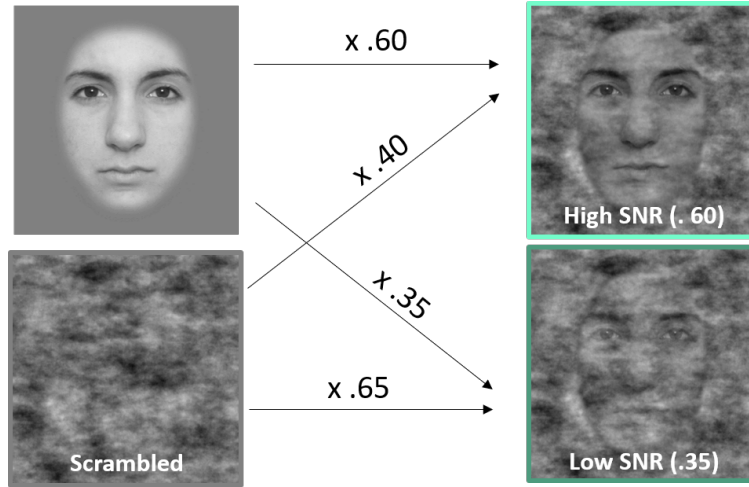


FIGURE 4.1: Alpha blending was employed to generate the varying levels of stimulus SNR. To seamlessly integrate the face with the phase-scrambled background, alpha blending, a technique based on pixel transparency, was utilised. Both the face image and the background underwent multiplication, each with its respective weight. For instance, a low SNR stimulus with a SNR of .35 was achieved by multiplying the face with a weight of .35 and the scrambled image with a weight of .65.

In order to minimise edge artefacts, we applied a face-shaped aperture to both stimuli and phase-scrambled masks. The aperture borders gradually transitioned from 1 to 0 to ensure a smooth transition. This resulted in the face embedded in the phase-scrambled background. The experiment was divided into 20 blocks per condition, with the 10 phase-scrambled backgrounds being repeated twice. In contrast to the stimuli, the masks were superimposed on the background without alpha blending to ensure strong masking. The final stimuli were cut out, resulting in a circular aperture with a diameter of 550 pixels (5.5° of visual angle).

Functional localiser

The stimuli employed for the localiser experiment were derived from established research and accessed through the fLoc functional localiser package (Stigliani et al., 2015). The three image categories - faces, hands, and instruments - each included 20 grayscale images, along with their corresponding phase-scrambled versions. The stimuli were composed of unfamiliar faces (half male) in varied viewpoints (including hair), isolated hands in diverse poses, and stringed instruments (such as guitar, cello, lute, etc.) presented in different orientations. The stimuli were superimposed onto Fourier phase-scrambled backgrounds, which were generated using the iterative scrambling procedure detailed above. As for the main experiment, the phase-scrambled backgrounds had a circular aperture with a diameter of 550 pixels (5.5° visual angle).

Procedure

All subjects completed a total of 20 experimental runs, which were conducted over two separate scanning sessions. Additionally, the first session encompassed a MR2RAGE anatomical scan as well as a T2-weighted anatomical scan, while the second session involved two functional face localising runs. Block designs were employed for both experimental and localiser functional runs, with 10-second blocks and alternating periods of 10 seconds of fixation. The runs also included a 12-second fixation period at the beginning and end. The fixation period was a uniform grey screen, with a fixation cross consisting of two diagonal thin black lines crossing at screen centre. Subjects were instructed to fixate their gaze on the central intersection of the cross throughout all runs. To maintain subjects' attention during functional runs, we implemented a task that required them to detect rare and brief colour changes in the stimulus. The subjects were instructed to press a button using their right index finger when they detected a colour change. In the main experimental runs, there were three red targets (HSV profile = 0.8, 1.0, 1.0) randomly distributed across the three equal parts of the run. For the functional localiser run, two blue targets (HSV profile 1.0, 1.0, 0.8) were presented, with one change per block half. It should be noted that no colour change occurred during the first or last stimulus of a block.

SNR test

During the first scans of the MRI scanner prior to the main experiment (known as the survey and shimming process - where the scanner is optimised to the individual to ensure the best possible image quality), subjects underwent an SNR thresholding task to identify the optimal SNR level for the main experiment. Subjects performed a gender classification task on face stimuli presented one by one for 200ms, followed by a 200ms mask, followed by a blank background until a response was recorded. Subjects were instructed to indicate the gender of the presented face by pressing the right index finger for male and the right middle finger for female. Four different SNRs (25, 30, 35, and 40) were used, with 40 trials per condition presented in a random order. The average accuracies achieved by the subjects at each SNR level were as follows: 54% (SD = 7.8), 64% (SD = 16.0), 66% (SD = 19.1), and 71% (SD = 14.9), respectively. For the main experiment, the low SNR condition was set at the SNR level corresponding to an 80% accuracy threshold. Specifically, 4 subjects were assigned a low SNR value of 30, while 6 subjects were assigned an SNR of 35, reflecting relative consistency in performance of subjects.

Main experiment

Main experiment runs lasted for 5-minutes. We administered 12 distinct experimental conditions - three stimulus types (low SNR, high SNR and scrambled) by four SOAs - alongside two retinotopic localiser conditions (Figure 4.2a). To counterbalance the order of experimental blocks, we randomised their presentation, while ensuring that no condition immediately followed any specific condition more than thrice over the course of the experiment. The two retinotopic localiser conditions, presented at the end of each run, entailed a single block of 4Hz contrast-reversing checkerboard stimuli. In one of these retinotopic blocks, the checkerboard was restricted to the stimulus region containing the face, whereas checkerboard covered the background region in the other block. This enabled us to localise the cortical area of the (extra)striate cortex that selectively responded to the stimulus region containing the face.

In each experimental block, 20 stimulus trials and four pseudo trials were presented (Figure 4.2b). Pseudo trials consisted of a blank screen with the background only and were randomly presented twice in the first half and twice in the second half of each block. The inclusion of pseudo trials was intended to increase unpredictability in the appearance of the stimuli. The remaining 20 trials contained stimuli of the same experimental condition, with one stimulus type (low SNR, high SNR, scrambled) and SOA per block. The order of the 20 stimulus trials was randomised with the constraint that each trial was equally likely to appear at the beginning, middle and end of the block. The background was filled with a circular phase-scrambled noise, which remained stable within each block. Each trial lasted for 417ms, which was equivalent to 50 frames at 8.33ms per frame.

During each stimulus trial (Figure 4.2c), a face blended with noise for a specific level of SNR appeared on the phase scrambled background a duration of 50, 75, 100, or 150ms (adapted from Goffaux et al., 2011; Schuurmans et al., 2023), after which a 166-millisecond mask was displayed, and subsequently, only the background remained. The duration of each event within a trial was a multiple of the frame rate (8.33ms).

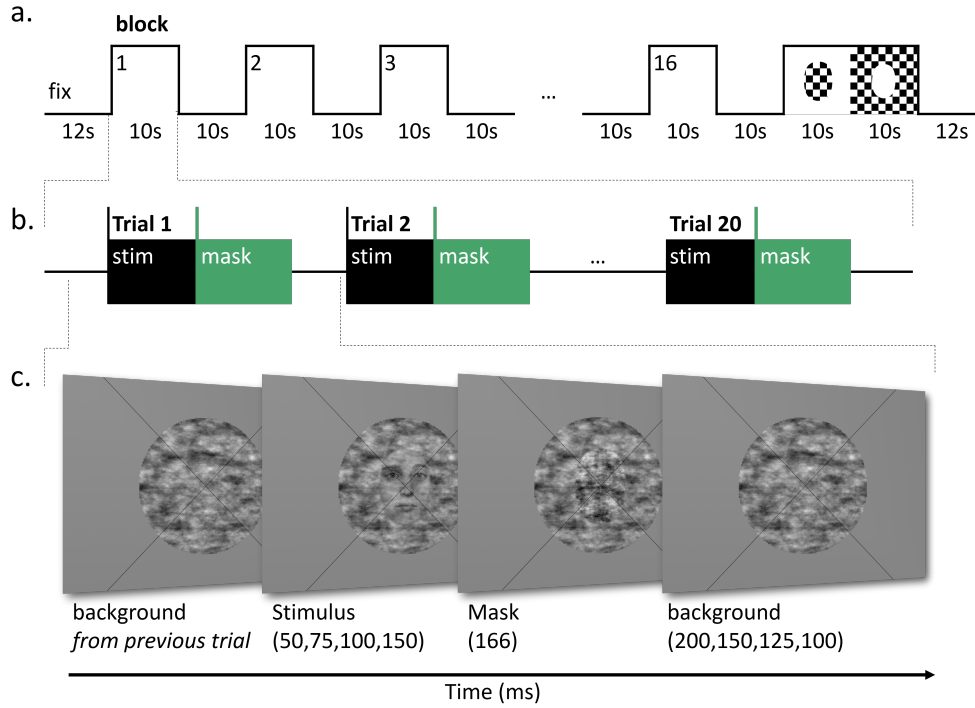


FIGURE 4.2: *Experimental design.* a) Each experimental run comprised a total of 16 blocks, which were alternated by fixation periods. Additionally, two retinotopic checkerboard blocks were administered at the end of each run. b) A block within the experiment encompassed 20 stimulus trials, all of which were from the same experimental condition (i.e. a specific combination of SNR and SOA). c) An exemplar trial: the presentation of a face stimulus, subsequently followed by a phase scrambled mask.

Functional localiser

During the two functional localiser runs lasting 6 minutes and 17 seconds each, subjects were presented with greyscale images of either intact or scrambled faces, hands, or instruments. Images from all categories were presented in a total of 36 blocks (6 blocks per condition), with each block corresponding to a single image category and containing ten images displayed in a random order. Each image was presented for 500ms, followed by a 500ms interstimulus interval. To ensure the localiser runs were similar (i.e. not all blocks of one condition in one run), both localiser runs contained 3 blocks per condition. The order of the blocks was counterbalanced across subjects, and no condition was repeated twice in a row.

MRI acquisition

MRI data was acquired using a 7T Philips Achieva MRI scanner (Philips Healthcare, Best, The Netherlands) equipped with an 8 channel transmit coil and a 32 channel receive coil (Nova Medical Inc, Wilmington, United States) at the Spinoza Centre for Neuroimaging in Amsterdam, the Netherlands. Universal pulses were employed for functional and anatomical imaging, as described in (Gras et al., 2017). Respiratory data was collected during brain imaging using a respiration belt, while cardiac data was collected using a pulse oximeter attached to a finger on the left hand.

As anatomical references, MP2RAGE (Marques et al., 2010) T1 anatomical MRI data were acquired at an isotropic spatial resolution of 0.7mm³, repetition time (TR) = 6.2ms, echo time (TE) = 2.35ms, flip angle (FA) = 6°, and field-of-view (FOV) = 245 x 245 x 184 mm. The second anatomical scan was a T2-weighted scan with the same voxel size and FOV as the MP2RAGE scan, but with a TR = 3s, TE = .93s, and FA = 90°. Functional data was acquired using a 3D-EPI sequence (Van Der Zwaag et al., 2012) which provides a higher temporal signal-to-noise ratio (tSNR) and enables a faster TR compared to a 2D sequence, therefore allowing to image the whole brain. The following sequence parameters were utilised: TR = 1.37 s, TE = 16.9ms, FA = 13°, voxel size = 1.79 x 1.79 x 1.8 mm, and FOV = 200 x 200 x 176 mm. Both magnitude and phase information were collected. To correct for distortions, we also obtained images using an identical sequence with phase-encoding in the opposite direction.

Preprocessing

The anatomical and functional data in this study were organised into the Brain Imaging Data Structure (BIDS; Gorgolewski et al., 2016) following conversion from the proprietary PAR/REC format to NIfTI using `dcm2niix` (v1.0.20220720). In future analyses, the physiological data will be formatted according to BIDS using `scanphyslog2bids`. Physiological noise correction will be performed using the RETROICOR method (Glover et al., 2000; Hutton et al., 2011), which employs Fourier expansions of varying order for the estimated phases of cardiac pulsation (3rd order), respiration (4th order), and cardio-respiratory interactions (1st order, A. K. Harvey et al., 2008). The corresponding confound regressors will be generated using the MATLAB PhysIO toolbox (Kasper et al., 2017). However, it should be noted that half of the subjects ($n = 5$) did not have successful physiological data collection.

Anatomical data

In this study, T1-weighted (T1w) and T1 map images were generated from the first and second inversion images of the MP2RAGE scan using the `Pymp2rage` software (v0.1). A sagittal sinus mask was created from the T2w image using the T1w/T2w ratio, and the subject's anatomy was registered to the MNI-space. The resulting warp files were used to transform the dilated sinus mask to subject space, and this was multiplied with the sinus mask derived from the T2w-image, excluding subcortical voxels. The resulting sagittal sinus mask was visually inspected and edited if necessary. Bias correction and denoising (using SPM12) were applied to the T1w image. We performed brain extraction and segmentations of white matter, grey matter, and cerebrospinal fluid using CAT12. Finally, all masks were combined and manually edited to create a mask that could accurately mask the sinus, dura, and tentorium. We conducted FreeSurfer (7.2.0) reconstruction prior to pre-processing using fMRIPrep 21.0.2 (Esteban et al., 2018, 2019), which is based on Nipype 1.5.0 (Gorgolewski et al., 2011, 2018). Next, within fMRIPrep, for each T1w image, non-uniformity intensity was corrected using the `N4BiasFieldCorrection` (Tustison et al., 2010), which was distributed with ANTs 2.3.3. This corrected T1w image was used as a reference throughout the pre-processing workflow. Subsequently, the T1w-reference was skull-stripped using the `antsBrainExtraction.sh` workflow (from ANTs) implemented in Nipype, with the `OASIS30ANTs` template as target. Volume-based spatial normalisation to MNI152NLin2009cAsym and MNI152NLin6Asym standard spaces was performed via nonlinear registration with `antsRegistration` (ANTs 2.3.3), using the brain-extracted versions of both T1w-reference and T1w-template.

Functional data

For each of the 20 functional runs per subject (across all tasks and sessions), we applied NOise Reduction with DIstribution Correction (NORDIC; Vizioli et al., 2020) outside of fMRIPrep to improve the SNR of the data. The subsequent pre-processing steps were performed using fMRIPrep. Firstly, a reference volume was generated and skull-stripped using a custom methodology of fMRIPrep. B0-nonuniformity maps (or fieldmaps) were estimated based on EPI references with a top-up (Andersson et al., 2003; FSL 6.0.5.1) approach. Prior to any spatiotemporal filtering, head-motion parameters with respect to the BOLD reference were estimated using transformation matrices, and six corresponding rotation and translation parameters (Jenkinson et al., 2002; mcflirt - FSL 6.0.5.1). The maximum movement on average was 0.70 mm (SD = 0.22 mm). The estimated fieldmap was then aligned with rigid-registration to the target EPI reference run and the field coefficients were mapped onto the reference EPI using the transform. The BOLD reference was subsequently co-registered to the T1w reference using boundary-based registration (bbrregister, FreeSurfer; Greve and Fischl, 2009), with six degrees of freedom (i.e. 3 rotations and 3 translations) configured for co-registration. Confounding time-series were calculated based on the pre-processed blood-oxygenation-level-dependent (BOLD) signals. Specifically, framewise displacement (FD), DVARS, and three region-wise global signals were computed. FD was computed by taking the absolute sum of relative motions (Power, Power et al., 2014) and the relative root mean square displacement between affines (Jenkinson, Jenkinson et al., 2002). Both FD and DVARS were calculated for each functional run, using their implementations in Nipype (Power et al., 2014). Motion outliers were identified by annotating frames that exceeded a threshold of 0.5 mm FD or 1.5 standardised DVARS. The three global signals were obtained from the cerebrospinal fluid (CSF), white matter (WM), and whole-brain masks. The functional data underwent spatial smoothing, using a Gaussian kernel of 5mm full width at half maximum, to compensate for inaccuracies in spatial registration and increase the SNR. It also underwent temporal smoothing with a high-pass filter, using a Gaussian-weighted least-squares straight-line fitting ($\sigma = 50.0s$), in FEAT (FMRI Expert Analysis Tool; Version 6.0; FMRIB's Software Library, www.fmrib.ox.ac.uk/fsl). A voxel-wise GLM was used to estimate each condition through a regressor that was set with block onset and a duration of 10 seconds, and convolved with a hemodynamic response function. The model included 15 regressors, consisting of 12 conditions, fixation, and 2 checkerboard conditions. To control for nuisance factors, 6 motion parameters and motion outliers, if present, were also included in the model as regressors. We generated contrasts for all 8 intact conditions by comparing them to the scrambled conditions. These contrasts were corrected for family-wise error and considered statistically significant at $p < 0.05$. In addition, we ran a second model without the contrasting the intact with scrambled conditions. Finally, we obtained GLM contrasted parameter estimates (i.e., beta-values) from the model.

Regions of interest

The delineation of V1/V2 in this study was achieved through a combination of anatomical landmarks and functional data. Initially, V1/V2 was anatomically localised using the Freesurfer atlas label. Subsequently, voxels exhibiting a significantly stronger response to the face-containing stimulus region compared to the background were selected following the methodology of S. A. Smith et al. (2010), de-Wit et al. (2012) and Schuurmans et al. (2023). To identify these voxels, the responses to the flickering checkerboard conditions in the final block of each run were contrasted: [stimulus - background]. The resulting significant voxels ($t > 1.65$, uncorrected $p < .05$) from this contrast, which fell within the anatomically defined V1/V2 region after smoothing, were designated as the V1/V2 ROI in this study (see Table 4.1 and Figure 4.3).

For each subject, the OFA and FFA were independently identified based on the functional localiser scan. Initially, brain regions that exhibited a higher responsiveness to faces than non-face objects were determined by calculating the conjunction of the [intact faces - intact hands] and [intact faces - intact instruments] contrasts. From these voxels, only those showing a significantly larger response to intact faces compared to scrambled faces [intact face - scrambled face] were selected to exclude activity driven solely by the amplitude spectrum of face images (Rossion et al., 2012; Schuurmans et al., 2023). The selection of significant voxel clusters on t-maps adhered to a $q[\text{FDR}] < .01$ criterion. Consistent with previous findings (e.g., Kanwisher et al., 1997; Rossion et al., 2000), a greater number of face-selective voxels were observed in the right FFA/OFA compared to the left (see Table 4.1 and Figure 4.3). The voxel coordinates (Table 4.1) aligned with prior research (Fahrenfort et al., 2012; Goffaux et al., 2011; Kanwisher et al., 1997; Rossion, Caldara, et al., 2003; Schuurmans et al., 2023; Thome et al., 2021).

TABLE 4.1: The average MNI152 coordinates and volume sizes of the ROIs across 9 subjects (with their corresponding standard deviations). The coordinates were derived based on the median voxel within each ROI.

ROI	Hemi	N voxels	X (L-R)	Y (A-P)	Z (S-I)
V1/V2	Left	439 (118)	-12	-75	-18
	Right	534 (153)	15	-73	-17
OFA	Left	624 (406)	-41	-66	-29
	Right	864 (705)	-43	-66	-28
FFA	Left	524 (487)	-39	-32	-39
	Right	759 (594)	42	-34	-39

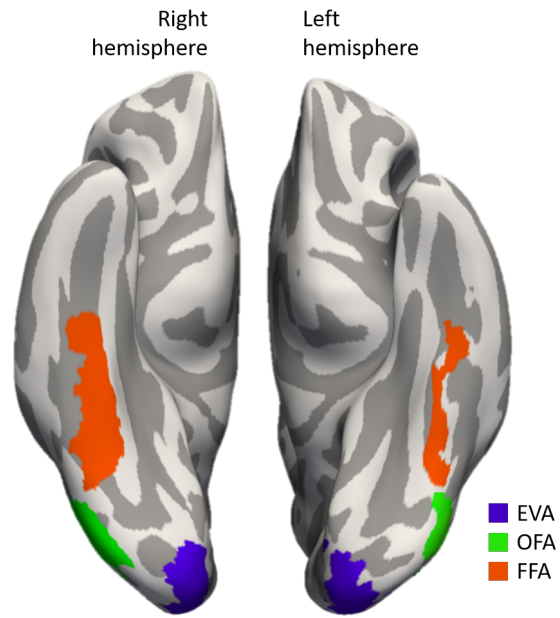


FIGURE 4.3: Functionally defined V1/V2, FFA and OFA - collapsed over 9 subjects - projected on the surface area of an inflated MNI152 brain. The ROIs are presented collectively for all subjects; nevertheless, it is essential to highlight that the analyses were conducted at the individual level, with individually defined ROIs.

Statistical analyses

Behavioural analysis

To ensure sustained attention to the stimulus during the course of the experiment, subjects were required to indicate whether the entire stimulus turned lightly red. Individual accuracy and response time (RT) in each condition were first submitted to a Linear Mixed effects Model (LMM; R). Next, a Type III Analysis of Variance (ANOVA) using Satterthwaite's approach was conducted to examine potential variations in accuracy and RT across conditions. In order to establish the most parsimonious model, we employed a stepwise variable selection procedure with backward elimination, following the methodology outlined by Heinze et al. (2018). The resulting model for accuracy included only a main effect of SOA (Bayesian Information Criterion [BIC] = -323.20):

$$betavalue \sim 1 + soa + (1|subject)$$

The final model for RT incorporated only main effects SOA and SNR (BIC = 54460.34). The model formulation for RT was:

$$betavalue \sim 1 + soa + snr + (1|subject)$$

Region of interest analysis

Individual beta weights for each condition of the main experiment were estimated using GLM from the average time course of each ROI (i.e. left and right V1/V2, OFA, and FFA). Initially, our plan was to compare the intact-scrambled contrast within each experimental condition to disentangle the activity associated with face stimulus processing from general visual responses. We anticipated stronger responses in the V1/V2 region for fully scrambled stimuli compared to intact stimuli (as found in studies de-Wit et al., 2012; Freud et al., 2017; Goffaux et al., 2016; Kok and de Lange, 2014; Markov et al., 2014; Olshausen and Field, 2006; Schuurmans et al., 2023). However, contrary to our expectations, we did not observe this effect. The responses to the scrambled condition were similar to those elicited by intact faces (Figure 4.9). After further reflection, the absence of intact-scrambled difference is not so surprising considering the important amount of noise that was contained in high and low SNR blocks. For the subsequent analysis, we focus exclusively on the intact low and high SNR conditions and exclude the scrambled data.

In the second-level analysis conducted in R, a Linear Mixed Model (LMM) approach was applied to fit the beta-values obtained from the first-level GLM for each ROI. The fixed effects in the model included SOA, SNR, hemisphere, and their interactions. To account for global variability among subjects, individual subjects were modelled as random effects, allowing for variations in intercept around the group mean. Although the LMM adeptly combines fixed and random effects to accommodate data variance, it is crucial to consider its inherent

constraints. Specifically, the model's reliance on fixed effects to account for variables such as SOA and stimulus type may not encompass the full spectrum of individual differences. Such a constraint could potentially limit the applicability of our findings to the general population. Model fitting involved the utilisation of a stepwise variable selection procedure, following backward elimination as described by Heinze et al. (2018). The initial model encompassed all variables and interaction terms (full model), which was also used for illustrative purposes when plotting the data. Subsequently, non-significant three-way and two-way interactions were sequentially removed to derive the simplest model that maximised the power of the analysis. The resulting model was validated by assessing whether it yielded the lowest BIC. The final model for V1/V2 with SOA as continuous factor, included SOA, SNR and hemisphere as main effects (BIC = 17695.48):

$$betavalue \sim 1 + soa + snr + hemi + (1|subject)$$

Given the trend of a potential interaction effect evident in the data we decided to further investigate by treating SOA as a discrete factor rather than continuous. This approach enabled us to examine the effects at specific timepoints and explore potential differences. By employing stepwise elimination, we arrived at a model that included a two-way interaction between SOA and SNR, as well as a main effect of hemisphere (BIC = 17744.87). The revised model formulation was:

$$betavalue \sim 1 + soa * snr + hemi + (1|subject)$$

As our expectations were based on linear increasing slopes in relation to SOA, we retained SOA as a continuous variable in the analyses for OFA and FFA. The final model for OFA incorporated the main effects of SOA (as continuous variable), SNR, and hemisphere (BIC = 9355.125). Similarly, the final model for FFA included only main effects (BIC = 19727.5). The model formulations for OFA and FFA were:

$$betavalue \sim 1 + soa + snr + hemi + (1|subject)$$

To interpret the estimated means, we utilised the "stats" package in R to conduct a Type III Analysis of Variance (ANOVA) employing Satterthwaite's method to determine the appropriate degrees of freedom.

Whole brain group-level analysis

To identify additional brain regions exhibiting enhanced recurrent processing for low SNR stimuli in comparison to high SNR stimuli, a group-level analysis covering the entire brain volume was conducted. Similarly to the previous analysis, a LMM approach was employed to fit the beta-values obtained from the first-level GLM. In this analysis, the same model was applied to each voxel in the fMRI data. Unlike the models utilised for the specific ROIs, all interaction terms were retained in the voxel-wise analysis, regardless of their significance. By maintaining consistency in the modelling across voxels, the interpretability of the statistical map would be preserved. Although this decision may have resulted in reduced power for lower-order interaction terms in the presence of redundant higher-order interaction terms, it was deemed essential for result interpretation considering that different voxels may have required the inclusion of distinct interaction terms. For every voxel, the complete LMM was fitted using the R software, incorporating all interaction effects:

$$betavalue \sim 1 + soa * snr + (1|subject)$$

Regrettably, no voxel exhibited significant interaction effects, and therefore, further interpretation of these results was not pursued. Detailed speculation regarding these findings is discussed in the discussion section.

Correlation analysis

In this exploratory analysis, we examined the inter-ROI functional relationships using two-sided Spearman's rho correlation analysis. It is important to note that such correlation analysis only provides an indirect measure of connectivity between brain regions. Our analysis focused on exploring the correlations between the beta values from separate conditions derived from the univariate GLM observed in the V1/V2, OFA, and FFA ROIs (collapsed over hemisphere). Each ROI was investigated separately for each subject across different experimental conditions. To account for multiple comparisons, we adjusted the significance level using an adjusted alpha level of .002. This adjustment was made considering that a total of 24 correlation coefficients were computed across the ROI pairs. By dividing the conventional alpha level of .05 by the number of correlation coefficients, we ensured a more stringent criterion for statistical significance.

Psychophysiological Interaction

In future investigations, we intend to explore the strength of coupling between the activity of V1/V2 and FFA or OFA through a psychophysiological interaction (PPI) analysis. PPI analysis incorporates the context of task-related neural activity and measures how the interaction between brain regions changes with task demands. By examining the task-dependent modulation of connectivity, PPI analysis provides a more direct and nuanced understanding of the functional interactions between brain regions. In the PPI analysis, the average time-series of the V1/V2 ROI will serve as the seed, extracted using the *fslmeants* tool to generate a physiological regressor. For each condition, a psychological regressor will be created by convolving a double gamma Hemodynamic Response Function (HRF) with the task regressor, following the approach described by O'Reilly et al. (2012). The PPI regressors will be constructed by multiplying the physiological and psychological regressors, generating one regressor per condition. Additionally, the model will incorporate the six motion regressors to account for motion-related artefacts. Subsequently, the GLM will be applied to all subjects, where the different regressors will be included, and the PPI interaction terms will be estimated specifically for the FFA and OFA regions. Next, similar to the main analysis, we will employ the stepwise variable selection procedure with backward elimination using LMM. The process will commence with the most complex model, progressively removing non-significant interaction terms, thereby identifying the simplest model that captures the data best.

Results

Behavioural results

During the main experiment, subjects completed in a colour change detection task. The colour of the stimulus changed randomly twice within each block. On average, subjects had an accuracy of 98.9% (SD = 1.01). Using an LMM with full interaction effects, we found that the subsequent ANOVA revealed a statistically significant main effect of SOA ($F = 2.681$, $p = .009$). The variability in performance attributed to the global differences between subjects accounted for a minimal portion of the overall residual variability (.49%). There was an increase in colour change detection accuracy as a function of SOA (Figure 4.4), from an average of 99% at the earliest SOA (50ms) to a perfect score at the longest SOA (150ms). However, this accuracy difference was minor. Given the simplicity of the task - designed primarily to maintain subjects' alertness and focus on the stimuli - performance appeared to be near the upper limit of response accuracy, indicating a potential ceiling effect. Regarding response times, subjects exhibited an average of 370.61ms (SD = 22.00ms). Employing the same LMM approach with full interaction effects and subsequent model pruning to retain only the main effects, the ANOVA indicated that there were no significant differences between conditions in response times ($ps > .130$).

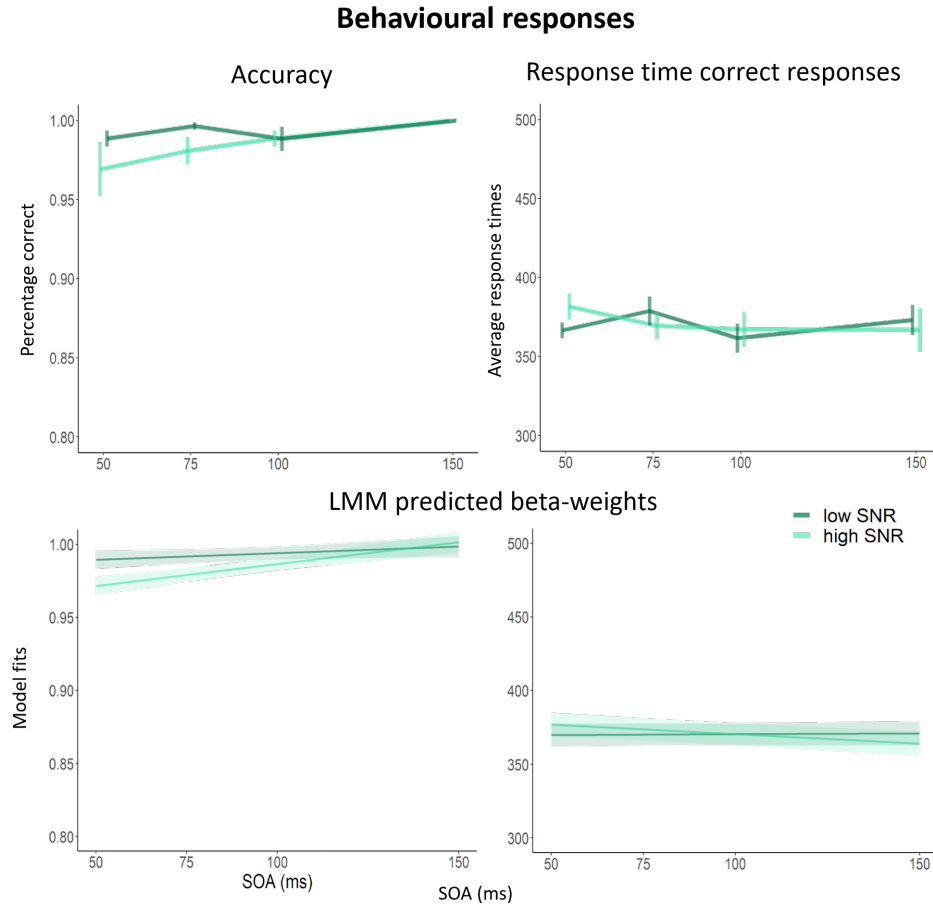


FIGURE 4.4: Results for response time of behavioural responses. The top row depicts the average group responses including a 95% within-subjects confidence interval Loftus and Masson (1994). On the bottom row the model predictions of the LMM are shown. The model for accuracy (bottom left graph) included the significant interaction of SOA with SNR. The model for response time (bottom right graph) included the significant factors of SOA and SNR. The graph was based on a model with the interaction effect for illustrative purposes.

Region of interest results

V1/V2

Based on our hypothesis, we expected an increase in recurrent processing in response to noisier stimuli within the V1/V2 regions. This elevation in recurrent processing would result in that longer SOAs allow more time for these processes to occur, subsequently leading to sparser representations and a reduced response. Specifically, we posited that low SNR stimuli engage more recurrent processing, would display stronger variations in response as a function of masking SOA. In contrast, high SNR stimuli, requiring less recurrent processing, would likely exhibit less variation in response to backward masking. In this context, we anticipated observing linearly declining slopes as a function of SOA for both low and high SNR conditions, with a more pronounced decline for the low SNR condition. Although our LMM analysis did reveal negative slopes for both conditions, the distinction between the SNR conditions was less clear than expected, as depicted in Figure 4.5.

Following the systematic removal of non-significant higher-order interaction terms through a Type III ANOVA, the final LMM consisted solely of main effects. The LMM revealed significant effects of SOA ($F = 3.882$, $p = .049$) and hemisphere ($F = 4.511$, $p = .034$). The global differences between subjects accounted for a small proportion of the residual variability (0.052%). Figure 4.5 displays the average BOLD activity and LMM predictions based on the full model, including terms that were not all statistically significant but were included for illustrative purposes. Although no significant findings were obtained for SNR, there appeared to be a potential trend in the data. Despite the non-significance of the SOA by SNR interaction term, there seemed to be a stronger negative slope as a function of SOA for low SNR stimuli (left hemisphere: $\beta = -.323$, $CI = .736$; right hemisphere: $\beta = -.265$, $CI = .737$), while high SNR masking resulted in a less negative slope (left hemisphere: $\beta = -.097$, $CI = .736$; right hemisphere: $\beta = -.055$, $CI = .737$). Nevertheless, these slopes did not significantly differ from each other (left hemisphere: $z = -.850$, $p = .395$; right hemisphere: $z = -.787$, $p = .431$).

Given these outcomes, and the unexpected non-linearity in the responses across SOAs, further investigation was conducted by treating SOA as a discrete rather than continuous variable. This approach aimed to discern finer differences in SNR conditions at each specifically tested SOA. In this discrete model, the random effect of subject accounted for a similarly small proportion of the residual variance (0.052%), mirroring the model where SOA was continuous. Notably, analysis revealed a significant interaction between SOA and SNR ($F = 3.281$, $p = .020$), supporting our hypothesis of temporal modulation for recurrent processes. Pairwise comparisons provided further insights, showing that there was a difference between SNR conditions, most pronounced at an SOA of 50ms ($z = 2.88$, $p = 0.003$). This finding corroborates our prediction about the highly disruptive effect of backwards masking on low SNR stimuli as compared to high SNR stimuli, which heavily rely on recurrent processing. This effect, however,

was not observed at other SOAs ($p_s > .140$). Additionally, a significant main effect of hemisphere was found ($F = 4.518$, $p = .034$), indicating lateralised differences in processing with overall higher responses in the left hemisphere. This may indicate a heightened sensitivity of the left V1 to backward masking of faces.

FFA

Similar to V1/V2, the random effect of subject accounted for a minimal proportion of the residual variance in the FFA (0.040%). The LMM applied to the FFA data incorporated main effects of SOA, SNR, and hemisphere. Contrary to the findings in V1/V2, all slopes observed in the FFA were positive (Figure 4.6), consistent with the region's known propensity to increase their response when input is more distinctly face-like and recognisable (Kanwisher et al., 1997; McCarthy et al., 1997; Rossion et al., 2012; Schiltz & Rossion, 2006; Yue et al., 2013). This is not only reflected in the significant main effect of SOA ($F = 21.694$, $p < .001$) but also in the main effect of SNR ($F = 43.179$, $p < .001$). The FFA exhibited elevated responses when presented with a high SNR, indicating a more robust representation of the face. Additionally, the main effect of hemisphere ($F = 156.178$, $p < .001$). The right hemisphere demonstrated higher responses compared to the left hemisphere, which aligns with the knowledge of the right hemisphere's tendency to exhibit greater responsivity to faces.

OFA

The random effect analysis revealed that the subject factor contributed minimally to the residual variance in the OFA region (0.061%). In line with the findings in the FFA region, all slopes observed in the OFA exhibited positive values, as illustrated in Figure 4.7. This pattern is consistent with the well-documented phenomenon of the OFA increasing its response to stimuli that are more perceived as face resembling (Kanwisher et al., 1997; McCarthy et al., 1997; Rossion et al., 2012; Schiltz & Rossion, 2006; Yue et al., 2013). This notion is supported by both the significant main effect of SOA ($F = 30.870$, $p < .001$) and the main effect of SNR ($F = 59.023$, $p < .001$). These findings suggest that the OFA region exhibited enhanced responses when presented with high SNR stimuli, indicative of a clearer and more discernible presentation of the face. Furthermore, similar to the FFA results, a main effect of hemisphere was observed ($F = 58.980$, $p < .001$). Specifically, the right hemisphere exhibited higher responses compared to the left hemisphere, aligning with the well-established right hemisphere's increased responsivity to facial stimuli.

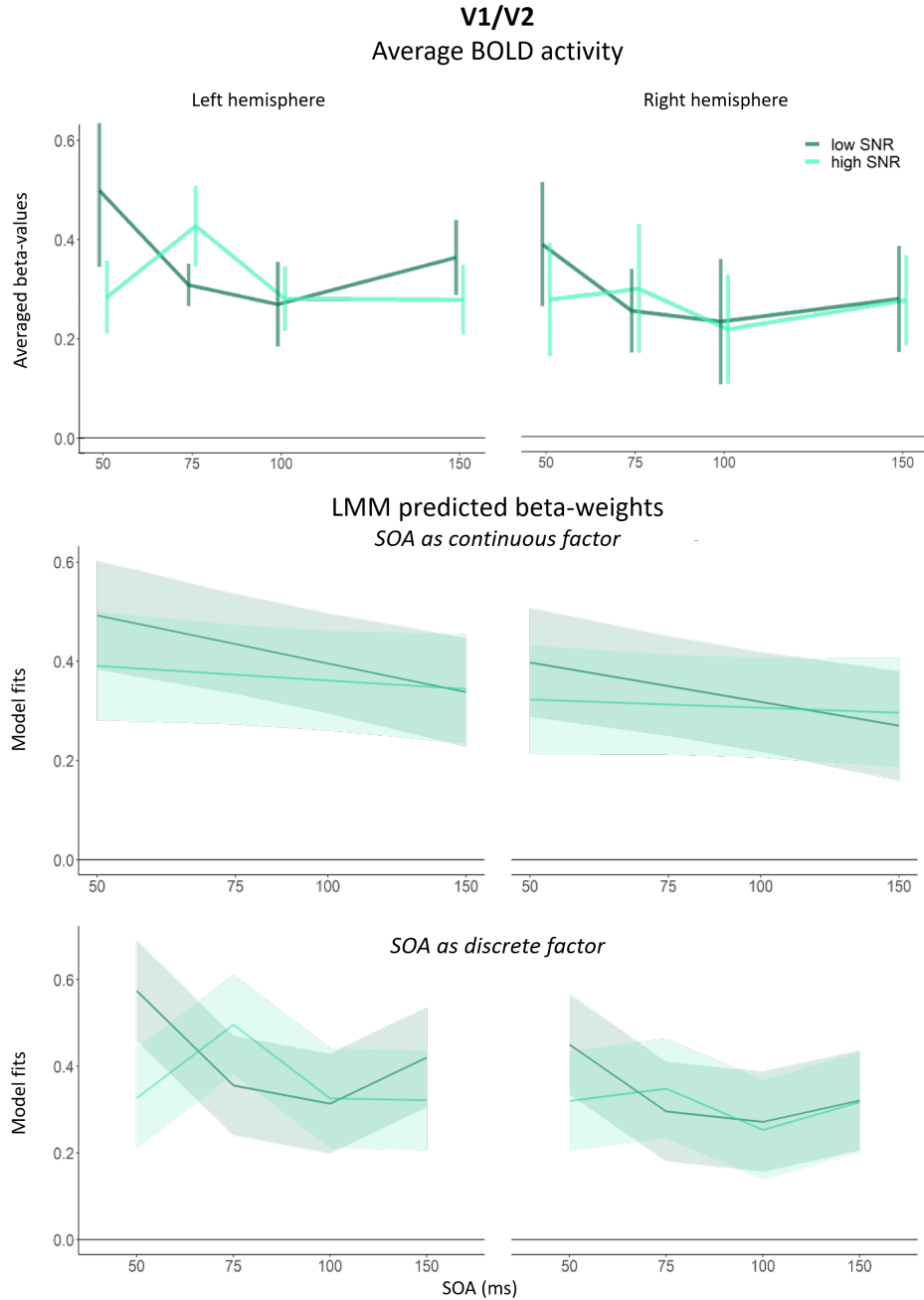


FIGURE 4.5: Left and right V1/V2 responses and model fits. The top row displays group-averaged beta-values, with 95% within-subjects confidence interval, following the methodology outlined by Loftus and Masson (1994). The LMM fits are presented in the two rows below. The middle row corresponds to the analysis treating SOA as a continuous factor. Despite observing significant main effects for SOA and hemisphere, the analysis did not yield significant findings for SNR, although a potential trend was discernible. Consequently, further exploration of the data was pursued to investigate this possibility. Hence, the bottom row represents the analysis treating SOA as a discrete factor. Notably, this model testing revealed a significant interaction effect between SOA and SNR, along with a significant main effect of hemisphere.

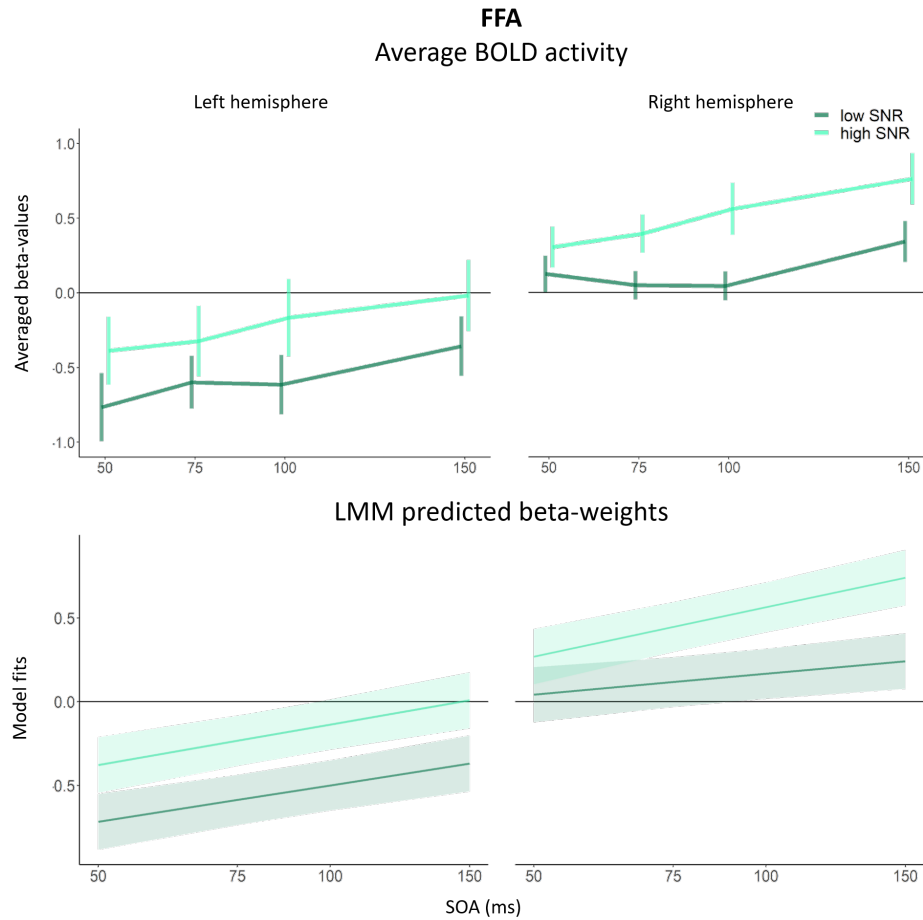


FIGURE 4.6: Responses in the FFA. The top row of plots presents the group-averaged beta-values, accompanied by a 95% within-subjects confidence interval, following the approach outlined by Loftus and Masson, 1994. The LMM fits shown in the bottom row correspond to a full model incorporating all interaction effects. It is important to note that this full model was employed solely for illustrative purposes. The statistical analysis revealed significant main effects of SOA, SNR, and hemisphere.

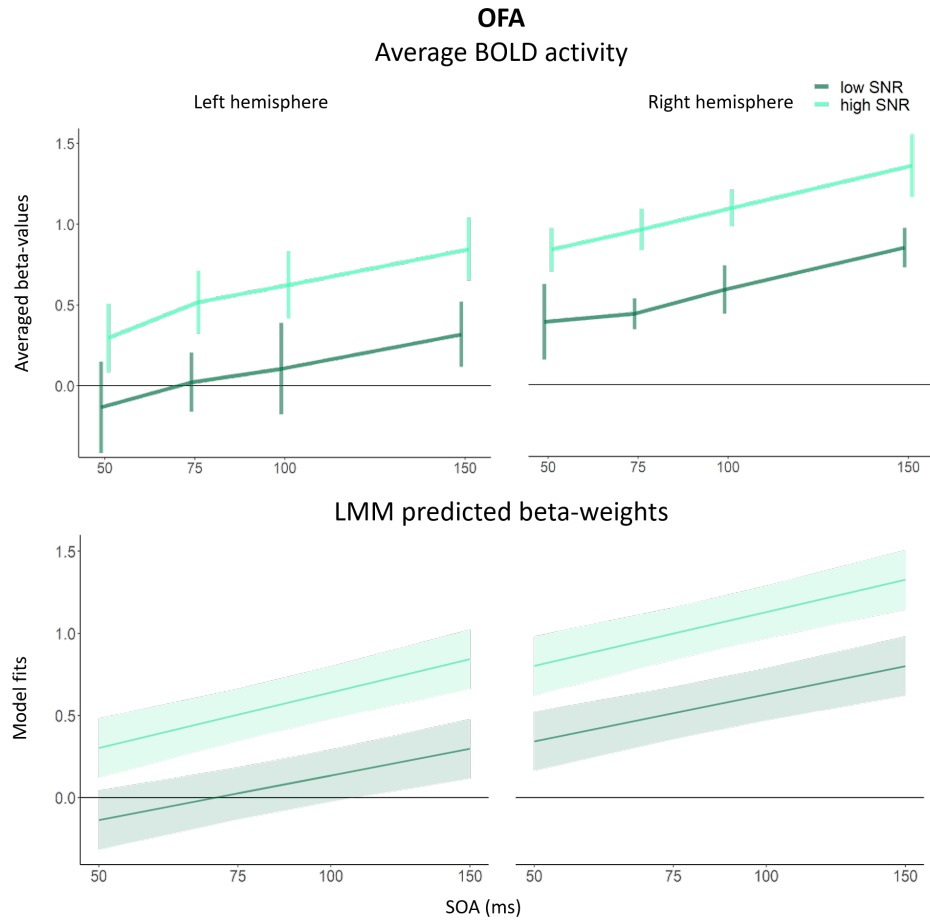


FIGURE 4.7: Responses and model fits in the OFA. The top row of plots displays the group-averaged beta-values, accompanied by a 95% within-subjects confidence interval, as per the methodology outlined by Loftus and Masson (1994). The bottom row presents the LMM fits obtained from a full model that included all interaction effects. The statistical analysis showed significant main effects of SOA, SNR, and hemisphere. Note: the full model was used purely for illustrative purposes.

Correlation results

In our exploratory correlation analysis, we probed the dynamic interactions between V1/V2, FFA, and OFA, across different levels of stimulus degradation and SOAs. Given the hypothesis that recurrent processing would increase when stimuli are more degraded, we expected to see negative correlations between V1/V2 and higher-level face selective regions that could reflect the feedback signals guiding V1/V2 activity, which in turn results in sparser representations and, therefore, weaker activity in these regions.

V1/V2 - FFA

In our study, negative correlations were consistently observed between V1/V2 and FFA across SNR conditions (see Figure 4.8 top row and Table 4.2). This pattern suggests that when the FFA responds less strongly to the presence of a face, particularly in low SNR conditions, there is increased activity in V1/V2. At low SNR, the correlation weakened as SOA increased ($\rho = -.259$, $p < .001$ at 50ms to $\rho = -.163$, $p < .001$ at 150ms), implying that other regions might become more involved in modulating V1/V2 responses over time. The similar trend was found for high SNR stimuli, with a negative correlation weakening in magnitude until 75ms ($\rho = -.160$, $p < .001$), hinting at reduced interdependence between V1/V2 and FFA, possibly because FFA resolves the face quickly and reduces reliance on V1/V2 for further processing.

V1/V2 - OFA

In contrast to the observations with FFA, the correlations between V1/V2 and OFA exhibited a generally positive pattern, strengthening with increasing SOA, particularly in the context of high SNR conditions (see Figure 4.8 middle row and Table 4.2). Examining low SNR stimuli, we found that the correlation between V1/V2 and OFA transitioned from being negligibly negative ($\rho = -.061$, $p = .055$) at 50ms SOA to weakly positive ($\rho = .115$, $p = .0003$) at 150ms SOA. This pattern reflects a shift in how these regions interact as SOA progresses, though the correlations remain relatively weak. For high SNR stimuli, the correlation became positive and strengthened consistently with increasing SOA ($\rho = .096$, $p = .003$ at 50ms to $\rho = .279$, $p < .001$ at 150ms). This significant positive correlation at high SNRs may suggest a more aligned processing pattern in both regions over time.

OFA - FFA

The observed correlation between OFA and FFA (Figure 4.8 bottom row and Table 4.2) exhibited a negative trend that strengthened with an increase in SOA for both low ($\rho = -.147$, $p < .001$ at 50ms to $\rho = -.216$, $p < .001$ at 150ms) and high SNR ($\rho = -.147$, $p < .001$ at 50ms to $-.201$, $p < .001$ at 150ms) conditions. This negative correlation is indicative of a possible competitive or divergent functioning between these two face-processing regions, each adapting differently to varying perceptual challenges. This finding is consistent with our understanding of both OFA and FFA as key components in face processing, interacting more dynamically when engaged with clear face stimuli.

TABLE 4.2: Spearman Rho coefficients and p -values of inter-ROI correlation for all conditions.

SNR	SOA	V1/V2 - FFA	p	V1/V2 - OFA	p	FFA - OFA	p
Low	50	-0.259	<.001	-0.061	.055	-0.147	<.001
	75	-0.235	<.001	-0.025	.440	-0.162	<.001
	100	-0.212	<.001	0.009	.781	-0.17	<.001
	150	-0.163	<.001	0.114	<.001	-0.216	<.001
High	50	-0.161	<.001	0.096	.003	-0.201	<.001
	75	-0.14	<.001	0.118	<.001	-0.219	<.001
	100	-0.042	.129	0.221	<.001	-0.219	<.001
	150	0.013	.650	0.279	<.001	-0.232	<.001

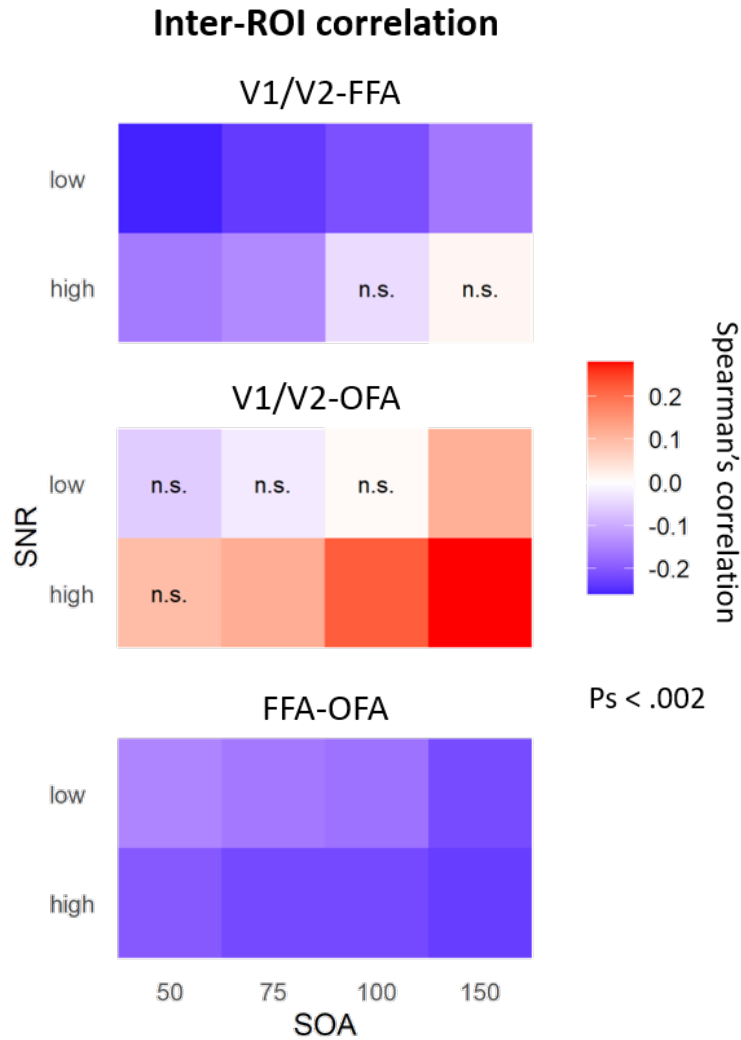


FIGURE 4.8: Spearman's rank correlation coefficients colour-coded for each ROI, and each condition. The figure presents the results of an inter-ROI correlation analysis investigating the functional relationships between the V1/V2, OFA, and FFA regions within the visual system. The analysis employed a two-sided test with an alpha level corrected for multiple analyses (adjusted alpha: 0.002). The colour scale represents the strength and direction of the correlations, with significant deviations from zero indicated by values outside the "n.s." (non-significant) range. The findings highlight the interconnectedness and functional dynamics between these regions for low-/ high SNR conditions and different SOAs.

Discussion

Recurrent processes play a pivotal role in the perception of visual input (Gilbert & Li, 2013; Hupé et al., 1998; Kravitz et al., 2013; Kreiman & Serre, 2020; Lamme et al., 1998; Petro et al., 2014), especially under degraded viewing conditions (Groen et al., 2018; Kar et al., 2019; Kreiman & Serre, 2020; Rajaei et al., 2019; Seijdel et al., 2020; Spoerer et al., 2018; Tang et al., 2018). However, the interplay between stimulus degradation and recurrent processing within the human visual system, in particular the role of early visual regions, remains unclear. Our research examined this dynamic by focusing on the involvement of early visual areas (V1/V2) and other visual regions such as the FFA and OFA in recurrent processing over time of face stimuli under varying conditions of stimulus degradation. Backward masking was utilised as a means to disrupt recurrent processes (Enns & Di Lollo, 2000; Lamme & Roelfsema, 2000) at different SOAs. We anticipated that with increased time for recurrent processes (later masking), V1/V2 would yield in sparser representations and reduced responses. Moreover, stimuli with high SNR, which are subject to swift processing and less recurrent processes, would be less impacted by backward masking. Conversely, stimuli with low SNR, requiring extended processing and greater recurrent involvement, were predicted to be more affected by backward masking. As a result, we anticipated witnessing declining V1/V2 responses with respect to SOA for both low and high SNR conditions, with a more pronounced masking modulation for the low SNR condition. Additionally, in face-selective regions (FFA and OFA), we expected increasing responses as a function of SOA, with stronger responses for stimuli of higher SNR, reflecting these regions' propensity to respond more to clearly defined faces.

In the V1/V2 regions, responses decreased as a function of SOA (Enns & Di Lollo, 2000; Lamme & Roelfsema, 2000). This pattern implies that as a stimulus - containing an organised shape - is processed over a longer time span, irrelevant aspects become suppressed (e.g. Kok and de Lange, 2014), sparsening its representation in V1/V2 (Olshausen & Field, 1996), and ultimately leading to overall lower BOLD responses (de-Wit et al., 2012; Freud et al., 2017; Goffaux et al., 2016). This decrease was expected to be steeper for low SNR stimuli, as they require more recurrent processing and are thus more susceptible to backward masking over time. Our LMM analysis in V1/V2 revealed negative slopes for both low and high SNR conditions, with little difference between the SNR conditions. However, when treating SOA as a discrete factor (i.e. analysing the data at specific SOAs) we found an interaction effect of SOA and SNR. Specifically, we observed a higher response for low than high SNR stimuli at early SOA, suggesting that the representation of low SNR input was initially less sparse than for high SNR input. This indicates that backward masking modulated responses for degraded stimuli stronger likely due to their higher reliance on recurrent processing for accurate visual perception, in line with behavioural findings by Wyatte et al. (2012). Furthermore, we examined the FFA and OFA regions and their responsiveness to different SNR stimuli over time. We consistently observed higher

responses in the right hemisphere compared to the left hemisphere in these regions in agreement with extensive body of research demonstrating that face-selective regions in the right hemisphere tend to be bigger and more responsive to face stimuli than those in the left (e.g., Kanwisher et al., 1997; McCarthy et al., 1997; Puce et al., 1996; Rossion et al., 2012; Sergent et al., 1992; Zhen et al., 2015). As expected, responses of both the FFA and OFA were enhanced for high compared to low SNR stimuli (e.g., Bankó et al., 2011; Heekeren et al., 2004; Hermann et al., 2015; Horowitz et al., 2004). In both regions, positive slopes were observed as a function of SOA, as longer exposure to face stimuli increased the responses in the face selective regions.

In our exploratory analysis, we aimed to test the relationship between responses of higher face-selective regions (OFA and FFA) and the V1/V2 responses at different levels of degradation over time. Specifically, we were interested in the changes in correlations that might reflect the high-level feedback signals guiding V1/V2 activity, which should manifest as negative correlations due to the top-down sparsening influence on V1/V2 response (e.g., Olshausen and Field, 1996; Schuurmans et al., 2023). Our results corroborated this hypothesis, as we found negative correlations between V1/V2 and FFA across all conditions. This negative relationship reflects that as FFA increases its response to faces, V1/V2 responses decreases; possibly due to the suppressive and sharpening feedback influence (O'Reilly et al., 2012; O'Reilly et al., 2013; Rajaei et al., 2019; Wyatte et al., 2012). This pattern was more pronounced in low SNR conditions, in agreement with past evidence that suggests stimulus degradation necessitates more extensive recurrent processing (e.g., Kar and DiCarlo, 2021; Kar et al., 2019; Kovács et al., 1995; Nielsen et al., 2006). Additionally, the negative correlation decreased as SOA increased suggesting that as more time is provided to process a face, the relationship between V1/V2 and FFA becomes less pronounced. One possible explanation for this weakened relationship might be that when FFA has clearly resolved a face, its interaction with lower visual regions such as V1/V2 decreases. Thus, the clarity of the face and the processing time appear to modulate the strength of the relationship between these regions. Interestingly, neural responses in V1/V2 and OFA were positively correlated, and these correlations intensified with SOA and were most pronounced under high SNR conditions. This pattern might suggest a more coordinated or similar processing between these regions as the face stimulus becomes clearer and more easily recognised. OFA is thought to process the face stimulus locally and that local processing relies on fine details (e.g., Goffaux et al., 2009, 2013), and therefore possibly requesting from V1 high spatial frequency information that is supposed to become available at later processing times (Peters et al., 2018; Schuurmans et al., 2023). We observed an overall negative correlation between FFA and OFA, with the correlation being stronger in high SNR stimuli and increasing in magnitude as a function of SOA. Previous studies have shed light on the divergent sensitivities of these regions, with FFA being more attuned to more global, and OFA local properties of a face (Andrews & Ewbank, 2004; Gilaie-Dotan & Malach, 2007; Goffaux et al., 2009, 2013; Levy et al., 2001). For example, Arcurio et al. (2012) found

that OFA responded maximally to isolated eye features but showed decreased responses when additional differing features, such as the nose and mouth, were presented in the display (see also Goffaux et al., 2013). This hints at a potential divergent functioning between these essential face-processing regions, each possibly adapting and responding differently to the perceptual challenges presented by stimulus degradation. However, it is important to note that correlations between regions do not necessarily point to direct influences from one region to another, such as those typically associated with recurrent processing. Instead, they may reflect simultaneous activation or parallel processing, rather than sequential or hierarchical interactions.

Our correlation analyses have shed some light on the interaction between V1/V2, OFA and FFA. However, it is important to recognise that they may not be the optimal approach to address functional connectivity. A more appropriate approach for studying connectivity, particularly in a blocked design such as the one used in the current study, would be the PPI analysis. The PPI analysis is structured around a regression model that integrates the interaction between a physiological variable of regional time series and a psychological variable of task design. Unlike the correlation of beta series method, PPI takes the time-series of a region into account within a GLM, enhancing the analytical power for block designs (Cisler et al., 2014). For these reasons, although the correlation analyses in the present study have yielded promising results, future investigations will benefit from applying PPI analysis to the current data. This method is likely to provide a more refined understanding of the connectivity patterns and is better suited for the design of this particular study.

Our analysis could be extended to encompass more regions that are recruited in recurrent processing of visual input. The inclusion of dorsal regions, for example, could provide further insights, as these areas are thought to represent shape prior to regions in the ventral pathway (Ayzenberg & Behrmann, 2022; Van Dromme et al., 2016). For instance, during the process of human face recognition, it is proposed that the posterior parietal cortex is responsible for representing the overall configuration of facial features and transmitting this information to face-selective ventral regions (Zachariou et al., 2017). Additionally, exploring the frontal regions could also provide valuable insights. For example, the orbitofrontal cortex is believed to quickly generate a probable interpretation of the incoming input (Ayzenberg & Behrmann, 2022; Ayzenberg et al., 2023; Bar et al., 2006; Bar, 2004; Fintzi & Mahon, 2014; Kauffmann, Bourgin, et al., 2015; Kauffmann, Chauvin, Pichat, & Peyrin, 2015; Kveraga et al., 2007; Mumford, 1992; Peyrin et al., 2010; Zachariou et al., 2017). Previous research demonstrated that the dorsolateral prefrontal cortex (DLPFC) enhances the processing of sensory input in the IT cortex when it is task relevant (e.g., FFA for faces, or PPA for scenes; Beck and Kastner, 2009; Gazzaley et al., 2005; Gilbert and Sigman, 2007; Miller and D'Esposito, 2005). DLPFC activation has also been observed when subjects mistakenly perceived degraded houses as faces, suggesting a role of frontal cortex in face pareidolia, as well as in face learning and retrieval (Devue et al., 2007;

Gobbini & Haxby, 2006, 2007; Sergerie et al., 2005; Sugiura et al., 2008). Moreover, evidence shows that feedback from the middle frontal gyrus to the FFA was amplified when subjects discerned whether degraded stimuli were faces or houses (Summerfield et al., 2006). Incorporating both dorsal and frontal regions into the investigation aligns well with the growing body of evidence suggesting that object recognition is not just a process localised in traditionally recognised areas in the visual cortex and ventral stream, but also critically involves both dorsal and frontal regions (Ayzenberg & Behrmann, 2022; Bar, 2004; Kay & Yeatman, 2016; Schuurmans et al., 2023; Zachariou et al., 2017). This comprehensive view will provide deeper understanding of the complex network that underpins our visual recognition capabilities.

We initially planned to use a whole-brain group-level analysis to investigate the involvement of dorsal, frontal, and other regions in recurrent processing during stimulus degradation. However, we did not identify any significant interactions or main effects in any of the voxels. Our utilisation of 7T fMRI, characterised by its robust SNR, warrants consideration with this respect. The heightened signal enabled by the high magnetic field strength results in overall smaller activated clusters or patches within subjects (Zhan et al., 2023). As a result, registration between different subjects can be challenging, as voxels with specific functions in one subject may not precisely overlap with those in another subject. For the whole-brain group-level analysis to be functional, we employed spatial smoothing in order to pinpoint voxels with similar sensitivities across subjects. Spatial smoothing however reduces the original signal from the targeted voxels while neighbouring non-activated voxels introduce disruptive noise (Scouten et al., 2006). This diminishes the effective signal in individual voxels, increasing the risk of false negatives. The implications are profound: spatial smoothing can render the detection of small amplitude activations practically impossible, even if they might be significant when assessed through more robust statistical methodologies (Stelzer et al., 2014). Considering the complexities of 7T fMRI data analysis, a more viable strategy may involve individually analysing each subject's responses. After this individual analysis, the patterns can be compared across subjects. Such a comparative analysis could involve examining common trends, variations in activation intensity or region, or exploring how individual differences might relate to variations in recurrent processing under different conditions of stimulus degradation. This approach mitigates the potential issues associated with spatial smoothing and aligns better with the strength of ultra-high field fMRI.

The expectations for our study were rooted in the anticipated differences in recurrent processing between high and low SNR stimuli. Specifically, we hypothesised that stimuli with high SNR, characterised by rapid processing and reduced reliance on recurrent mechanisms, would exhibit diminished susceptibility to backward masking as a function of SOA. On the other hand, stimuli with low SNR, demanding extended processing and greater recurrent involvement, were expected to be more disrupted by masking, particularly at short SOAs where most recurrent processes would be hindered. Consequently, we predicted

a steeper negative slope in the relationship between V1/V2 response and SOA for low SNR conditions compared to high SNR conditions. The LMM analysis, however, only revealed a trend in this direction, potentially due to a non-linear nature of the relationship between responses in V1/V2 and SOA. To investigate this trend further, we therefore decided to treat SOA as a discrete rather than continuous variable, thereby enabling an analysis based on separate SOAs. We indeed found a significant interaction between SNR and SOA. This effect was mainly pulled by a difference at the shortest SOA; higher responses for low SNR compared to high SNR conditions. This finding supports our initial hypothesis that backward masking would have a more disruptive effect on more degraded stimuli. However, the non-linear trend could indicate more complex dynamics in the recurrent processing of visual stimuli.

The initial design of our study incorporated a phase-scrambled condition as a control to enable the comparison of responses between phase-aligned and phase-scrambled face stimuli. This comparison was intended to isolate the visual responses unique to face shape processing (Fahrenfort et al., 2007; Goffaux et al., 2011; Grill-Spector et al., 2000). By maintaining the amplitude spectrum of the original image and disrupting the stimulus structure, phase scrambling produces unrecognisable images, thus making a suitable control of the physical properties of the stimulus. Typically, V1 activity is known to increase with stimulus scrambling in unconstrained viewing conditions, since there is less strong feedback to sharpen the representations (Freud et al., 2017; Goffaux et al., 2016; Lerner et al., 2001). In our experiment, we did not observe the expected higher response to scrambled than intact condition (see supplementary Figure 4.9), making the approach problematic and leading us to exclude the scrambled condition from our analysis. A probable explanation could be that across conditions stimuli were not only masked with noise, but were also embedded in phase-scrambled noise to various extent (moderate in high SNR, high in low SNR, and maximum in phase-scrambling condition). As a consequence, any given block contained a massive amount of noise, which might have reduced the differences across conditions, and in comparably strong responses for both intact and fully scrambled blocks (supplementary Figure 4.9). Another plausible reason is a phenomenon called face pareidolia (J. Liu et al., 2014; M. L. Smith et al., 2012; Summerfield et al., 2006; H. Zhang et al., 2008). The subject, repeatedly exposed to faces embedded in noise throughout the experiment, may have perceived faces even in the phase-scrambled noise due to its spectral properties being similar to that of the faces (Hansen et al., 2010). An fMRI study by H. Zhang et al. (2008) identified occipital regions showing greater activations when faces were perceived in noise, though they did not specifically report on V1 or V2 involvement. The overall noisy stimuli may have led to pareidolia in the fully scrambled conditions, therefore attenuating the intact/scrambled response differential we expected to observe in V1/V2. However, it is crucial to acknowledge that these speculations warrant further investigation to establish a more comprehensive understanding of the underlying mechanisms.

In conclusion, our study provides insights into the temporal dynamics of recurrent processing under different conditions of stimulus degradation. Our results contribute to understanding of the adaptability of the visual system, particularly in the context of processing complex and noisy stimuli such as (degraded) faces. V1/V2 responses decrease with longer exposure to stimuli, indicating sparsening of stimulus representation over time. Stimulus degradation increased V1/V2 susceptibility to backward masking disruption likely due to increased recurrent processing. The unique interactions observed among these regions further highlight the complexity of the mechanisms employed by the visual system to ensure accurate perception under challenging viewing conditions. Future investigations focusing on the coupling strength of these regions during enhanced or non-enhanced recurrent processing will pave the way for a more nuanced understanding of the brain's visual processing network.

Supplementary materials

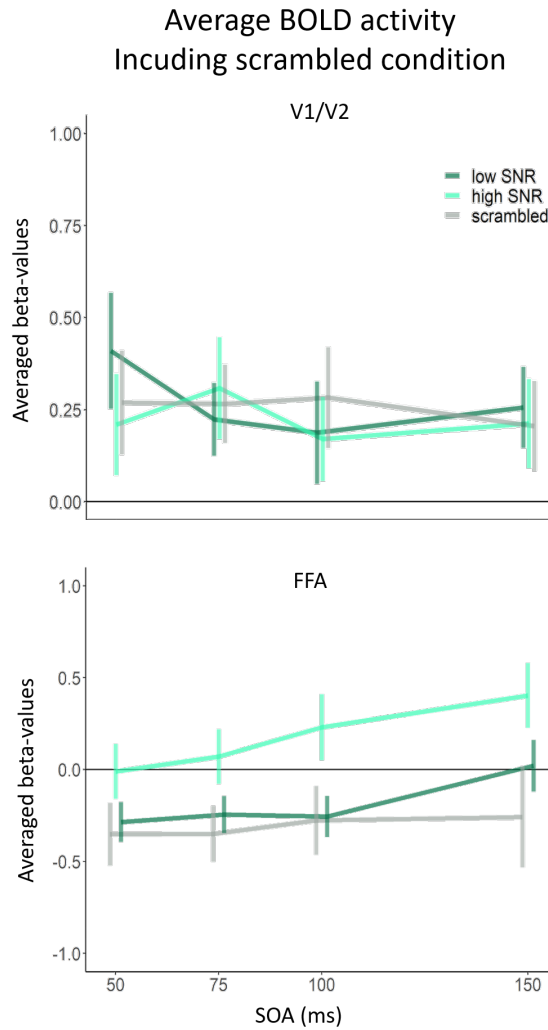


FIGURE 4.9: Responses in V1/V2 (top) and FFA (bottom) including the scrambled condition. The figures illustrate the group-averaged beta-values, complemented by a 95% within-subjects confidence interval, in accordance with the method delineated by Loftus and Masson (1994). Contrary to our expectations that the scrambled condition (represented by the grey line) would induce stronger responses than the intact ones (Freud et al., 2017; Goffaux et al., 2016; Lerner et al., 2001), the data did not reveal such a distinction.

Chapter 5

General discussion

Summary

This thesis is oriented towards enriching our understanding of the mechanisms underlying real-world visual perception. Central to this investigation is the exploration of recurrent processing dynamics within the human visual system. We employed backward masking, with a specific emphasis on face stimuli, to elucidate the role of V1 in recurrent processing. Our approach was guided by two queries concerning recurrent processing: first, the proposal that it evolves in a coarse-to-fine manner; and second, the proposition that recurrent processes increase as a response to stimulus degradation.

Chapter 2 and 3 addressed the proposed mechanism of recurrent processing, postulating that it optimises the processing of visual input in a coarse-to-fine manner. The theory suggests that coarse, or low spatial frequencies (LSFs), are processed initially, guiding the subsequent processing of finer, high spatial frequency (HSF) information. Within this paradigm, the primary visual cortex (V1) has been conceptualised as an 'active blackboard' that assimilates visual information in a coarse-to-fine manner (Bullier, 2001). Despite V1's presumed pivotal role in integrating complex visual stimuli, prior research had not thoroughly explored its exact function in the coarse-to-fine framework. Addressing this knowledge gap, our study presented in chapter 2 employed backward masking across different SF ranges in facial stimuli to disrupt either LSF or HSF processing. We hypothesised that if V1 indeed employs a coarse-to-fine processing strategy, its disruption should result in suboptimal representations within this region. Our findings revealed a particularly disruptive impact of early masking on LSF processing, diminishing over time, while HSF masking had increased disruptive effects at later stages. These effects were observed not only in V1 but also in ventral (fusiform face area, FFA), dorsal, and orbitofrontal brain regions. By maintaining a key aspect of naturalistic vision, specifically the full range of aligned SFs, our study extends beyond prior evidence of coarse-to-fine precedence by demonstrating that the integration of visual information in V1 operates in a coarse-to-fine manner. This postulated the question of how such disruption of coarse and fine scales would impact behaviour. Consequently, we further investigated the behavioural repercussions of disrupting coarse-to-fine processing in a face identity recognition task using familiar faces (Chapter 3). We predicted that LSF processing would be fast, transient, and less vulnerable to backward

masking, whereas HSF processing would be slower, more sustained, and dependent on recurrent processes. While part of our predictions was confirmed, later masking - which allows more time for recurrent processes - proved less disruptive for identity recognition. Hence, no effects from the SF manipulation of the mask were detected. This outcome poses questions, especially considering the significant masking effects on different SF ranges observed in our fMRI study. It is plausible that the coarse-to-fine process transpires within a brief time window, possibly too short to be captured in behavioural responses. This topic will be explored further in the subsequent discussion.

The thesis's second query regarding recurrent processing poses that there is an increase in recurrent processes as visual input is more degraded. Prior studies, using computer modelling, monkey physiology or human behavioural data, have indicated an increase in recurrent processes when a stimulus is noisy or degraded. The nature of this interaction between stimulus degradation and recurrent processing within early visual regions of the human visual system, however, remained unclear. We examined the contributions of early (V1/V2) and higher-order visual regions (occipital face area, OFA; FFA) during recurrent processing amidst stimulus degradation (Chapter 4). Employing 7T fMRI and backward masking, we studied the disruptive effects of masking on human face stimuli, adjusting both the stimulus onset asynchronies (SOAs) and signal-to-noise ratios (SNRs). Our results demonstrated that in V1/V2, for both levels of stimulus degradation, masking at later SOAs was less disruptive compared to earlier masking. This supports the notion that when there is more time for recurrent processes, irrelevant aspects of a stimulus become suppressed (e.g. Kok and de Lange, 2014), sparsening its representation in V1/V2 (Olshausen & Field, 1996), and ultimately leading to overall lower neural responses (de-Wit et al., 2012; Freud et al., 2017; Goffaux et al., 2016). This trend was especially clear with more degraded stimuli; bigger disruptive effects of (early) masking, presumably due to more recurrent processing. In the higher-order regions, OFA and FFA, we found an expected enhancement in the responses to clear face stimuli compared to those to more degraded stimuli (e.g., Bankó et al., 2011; Heekeren et al., 2004; Hermann et al., 2015; Horowitz et al., 2004). Moreover, both regions demonstrated an increase in their responses with longer SOAs, indicating that prolonged exposure to facial stimuli increased the responses within these face-selective areas.

In short, the studies in this thesis demonstrate the role of recurrent processing during the processing of complex (face) stimuli in various brain regions and its impact on behaviour. It highlights the mechanics of coarse-to-fine processing and the effects of different levels of stimulus degradation on recurrent processes. The forthcoming sections will delve deeper into nuanced aspects of the visual system that were uncovered through the research. This includes a closer examination of coarse-to-fine processing and why behavioural evidence might be elusive. The temporal dynamics within the visual system and the special emphasis on face stimuli will be explored. Attention will also be given to the paradigms of backward masking and the specificity and validity of the results, reflecting on how this method specifically interacts with recurrent processing. The choice of complex stimuli, such as faces, and its implications on SFs will also be evaluated. This discussion seeks to synthesise the findings, critically evaluate the methodology, and place the research within the broader context of understanding recurrent processes in visual perception.

Studying coarse-to-fine integration with full-spectrum stimuli

Our studies took an innovative approach to unravel the mechanisms of coarse-to-fine integration by manipulating the properties of the backward mask. By focusing on manipulating the backward mask, the integrity of the initial facial stimulus was preserved and closer to natural viewing conditions. In natural images low and high SFs are typically aligned in a way that the location of LSF features provides valuable information about the content and exact location of HSF contrast (Figure 5.1). This alignment of SFs underlies efficient visual perception (Bex & Makous, 2002; MaBouDi et al., 2016). It is postulated by efficient coding theories (H. Barlow, 2001; Simoncelli, 2003) that the visual system strives to maximise efficient processing by exploiting redundancies across image scales (Petrás et al., 2019, 2021). This optimisation is proposed to take the form of statistical dependencies across SF scales. This proposition aligns with the coarse-to-fine framework (Bar, 2004; Bullier, 2001), which is corroborated by the investigation presented in chapter 2 of this thesis (Schuurmans et al., 2023). In other words, the fast processing of LSFs in high-level areas of the brain is sent back via feedback signals to limit low-level processing to the most recognition-relevant HSF details, thereby sparing computational resources at the low level that would otherwise be devoted to HSF processing.

The coarse-to-fine model of visual processing is not only mechanistically efficient but also developmentally plausible. Infants, who initially perceive the world with low acuity, are attuned primarily to LSFs. As their visual system matures, their acuity improves, allowing for the perception of HSFs. Thus, the developmental trajectory of human vision aligns with the proposed mechanism, where a transition from an LSF-dominated view in infancy gradually gives way to the inclusion of finer HSF details, paralleling the brain's natural progression in processing visual information.

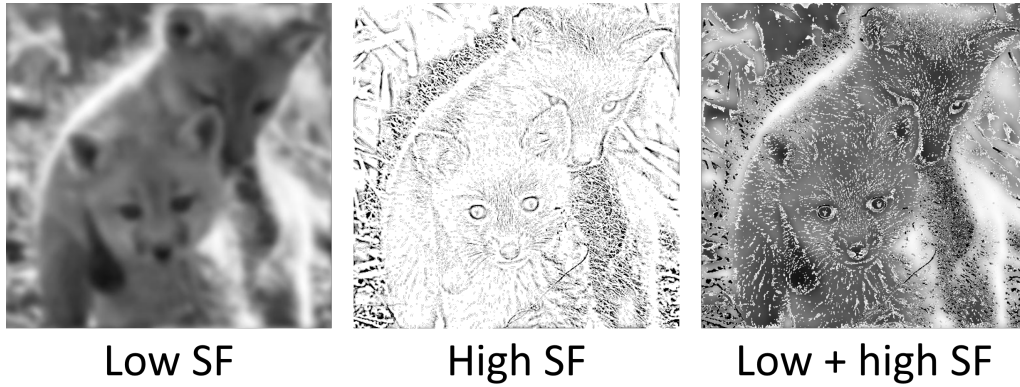


FIGURE 5.1: Natural images exhibit interrelated spatial frequency scales. The left image showcases the LSFs of baby foxes, while the centre highlights their HSFs. Rather than being entirely distinct, the information within these scales is interdependent; the low spatial frequencies provide cues about the locations of high spatial frequencies, as showcased in the right image. Such overlapping information between frequency scales enhances efficient visual perception due to inherent redundancy.

While visual input contains a broad range of SFs, prior research has mostly focused on examining coarse-to-fine processing by comparing neural responses to SF filtered stimuli that focus on narrow bands (e.g. De Cesarei and Codispoti, 2013; Goffaux et al., 2011; Musel et al., 2014). This approach, while providing valuable insights into coarse versus fine precedence, falls short when it comes to understanding coarse-to-fine integration. It often presents LSF and HSF separately in both space and time. Such an approach does not accurately reflect the correlated nature of LSF and HSF signals in natural images. This gap in representation led Petras et al. (2019, 2021) to examine the distinct influences of LSF and HSF processing on the human perception of SF aligned broad-band images of faces. By distinguishing SF at the cortical response level, rather than the stimulus level, the inherent characteristics of naturalistic broad-band vision are retained. This allows for a more ecologically valid exploration of coarse-to-fine integration. Employing electroencephalography (EEG), Petras et al. (2019) captured the neural responses associated with LSF and HSF in phase scrambled stimuli. These responses then served as training data for linear classifiers, which were subsequently assessed for their adaptability over a spectrum of broad-band visual stimuli. The classifier's role was to categorise the broad-band response as either LSF or HSF, based on a predominant presence of EEG scalp-patterns similar to processing of either LSF or HSF stimuli. Their findings suggest a dominance of LSFs in facial image representation preceding that of HSFs. Subsequent research using MEG, Petras et al. (2021) reinforced the idea that early LSF encoding can reduce the computational demands of processing HSF details, in congruence with coarse-to-fine theories. However, the studies by Petras et al. (2019, 2021) were not without limitations. The impact of contrast on visual processing is well-documented (e.g., Boynton et al., 1996; R. L. De Valois and

De Valois, 1980), and a typical method to balance the contrast of stimuli with different SF ranges in studies involves increasing the contrast of HSF images after filtering (Goffaux et al., 2011; Kauffmann, Chauvin, Guyader, & Peyrin, 2015; Mu & Li, 2013; Vlamings et al., 2009). To sidestep the resulting unnatural amplitude spectra, Petras et al. (2019) adopted a different approach. They designed 5th order Butterworth frequency filters with specific cut-offs that equalised the Fourier-space amplitudes of LSF and HSF images, averaged across all images. This was a critical step as lower SFs typically contain more energy, and without this adjustment, LSF images would be more dominant in full-spectrum stimuli. To avoid this disproportion, the researchers excluded the lowest 5cpf before applying the filters. Following this, the Fourier amplitude of each image was modified with these filters and converted back into the image domain. Broad-band images were then generated by merging the Fourier amplitudes of the original images with the sum of both low (5-8cpf) and high (>8cpf) SF filters before re-converting them into image space. This method effectively negated the need for post-filtering contrast enhancement of HSF images. Nevertheless, this technique had a drawback: it overlooked the extremely low SF ranges, which are crucial in face processing (Goffaux et al., 2003, 2005; Goffaux & Rossion, 2006; Quek et al., 2018; Vlamings et al., 2009). Motivated by these limitations, our investigations (detailed in chapters 2 and 3 aimed to probe coarse-to-fine processing within a more ecologically representative stimulus.

By backward masking full-spectrum stimuli, we were able to circumvent the problem of spectral power differences between SF filtered stimuli. Since we only studied the contrasted neural (blood-oxygen-level dependent, BOLD) responses related to the initial full-spectrum face stimulus, but not to SF filtered stimuli, the variations in energy across different SF ranges were not problematic. Our inspiration to use masking of various spatial scales was drawn from the study conducted by Bacon-Macé et al. (2005). In their study, white noise images were filtered at different spatial scales to create high-contrast binary patterns. These were used to mask a variety of image categories, and subjects had to indicate when they saw an animal (categorisation task). The results showed that the efficacy of masking was largely dependent on the spatial scale of the mask. At short SOAs, when masks with the finest spatial scales were used, there was a strong disruptive effect on categorisation performance. At coarser scales, however, the masking was less effective. In line with this, we also opted for masking of spatial scales and utilised low and high SF filtered masks. This novel approach to study coarse-to-fine processing proved fruitful as the robust results point out in our study described in chapter 2.

V1 as an "active blackboard"

Our research challenges the convention that V1 mainly serves as an early filter for visual information relayed to high-level regions (Kravitz et al., 2013). We provided evidence supporting the idea that V1 integrates bottom-up input with top-down influences from ventral, orbitofrontal and dorsal areas (e.g., Alink et al., 2010; Astorga et al., 2022; Ayzenshtat et al., 2012; Chadwick et al., 2023; Kazandjian et al., 2021; Koivisto et al., 2011; Kok and de Lange, 2014; Lamme, 1995; Lamme et al., 1998; Muckli, 2010; Noudoost and Moore, 2011; Rademaker et al., 2019; Roland et al., 2006; Super et al., 2001; Zipser et al., 1996). This suggests V1's involvement in more advanced visual functions than typically presumed. The coarse-to-fine dynamics we observed in V1 during processing of complex face stimuli reinforce the idea of the V1 functioning as an 'active blackboard' for recurrent refinement of visual representations (Bullier, 2001; Kok & de Lange, 2014; T. S. Lee et al., 1998; Mumford, 1992).

These observations seem to diverge from the coarse-to-fine model posited by Bar et al. (2006) and Bar (2004). Their model suggests that LSF information rapidly reaches the prefrontal cortex (pFC), prompting potential semantic interpretations of the stimulus based on its coarse properties (see Figure 5.2). These predictions are then relayed back to the inferotemporal (IT) cortex through feedback connections, guiding and facilitating the subsequent processing of the HSF information. Notably, this model side-lines V1, depicting it merely as a "relay station", with the real integration of SFs occurring in the IT cortex. Contrary to this, our findings underscore V1's active involvement in the coarse-to-fine integration of visual data. This lends greater credence to the alternative proposed by Bullier (2001), which envisions V1/V2 as an 'active blackboard' for SF integration.

Investigating recurrent processing

The experimental designs throughout this thesis were selected to examine recurrent processing, particularly when faced with varying conditions such as alterations in SF access and under differing degrees of stimulus degradation. Specifically, we designed our experiments to minimise the possibility that the observed modulations of brain responses arise solely due to differences in feedforward processing.

First of all, we employed full-spectrum stimuli across all conditions and we made sure these full-spectrum stimuli were uniformity in the SF, luminance, and contrast of the stimuli across all conditions. Additionally, in the studies from chapters 2 and 4, a phase-scrambled control was incorporated, wherein the original image's amplitude spectrum was retained, but its structural coherence was scrambled, producing unrecognisable images. By ensuring that both intact and scrambled stimuli were identically masked, we were able to effectively subtract the mask's direct impact on the BOLD response, as reflected in prior studies (Fahrenfort et al., 2007; Scholte et al., 2008). This intact-scrambled subtraction method also accounted for the mask's influence on basic stimulus attributes,

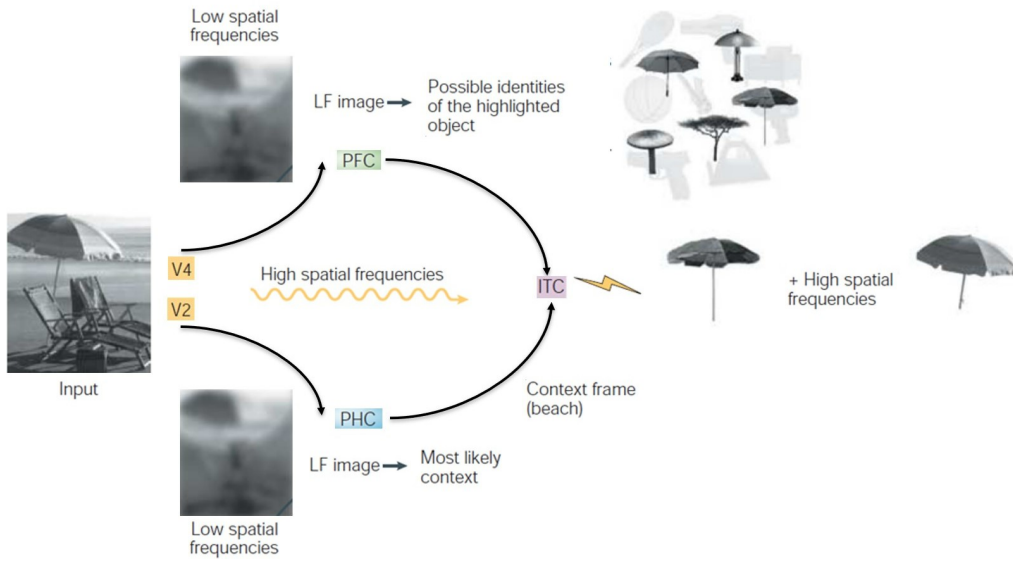


FIGURE 5.2: A model proposing contextual facilitation in object recognition; figure adapted from Bar (2004). The initial conjunction of the associated context and the possible interpretations of the individual target object based on LSFs enables swift activation of the representation of a beach umbrella. This early representation is then refined based on the later arrival of higher SFs. In this illustration, only the relevant connections and flow directions are depicted for simplicity. In this model by Bar (2004) V1/V2 only relay visual information, it is important to note that this is not in line with the proposed model by Bullier (2001) (Figure 1.5), who proposes V1/V2 (instead of ITC) as an "active blackboard" important for the integration of HSF input. ITC refers to the inferior temporal cortex, PFC represents the prefrontal cortex, and V2 and V4 represent early visual areas. LF denotes low frequency and the symbol of a 'lightning strike' indicates activation of representations.

such as orientation or SF, neutralising the direct effects of mechanisms attuned to these properties. An MEG study by Petras et al. (2021) further showed that phase scrambling does not alter the initial feedforward responses in areas such as V1/V2. Consequently, if scrambling the images does not affect the initial feedforward responses, then any observed differences in response amplitudes between intact and scrambled conditions in our study likely result from feedback or lateral interactions, which are characteristic of recurrent processing. This insight lends weight to the interpretation that the differential response we examined (in chapter 2) relates to recurrent processing. However, in the 7T fMRI (Chapter 4), the scrambled condition did not yield the anticipated results. This discrepancy nudged us to exclude the scrambled condition from our analysis and focus on the uncontrasted (i.e., no intact-scrambled subtraction) conditions instead.

A crucial assumption in the studies of this thesis is that backward masking disruptions recurrent processing. However, our experiment was not specifically designed to discount a feedforward hypothesis. The hypothesis that backward masking operates in a feedforward manner suggests that the mask's feedforward signal interferes with the processing of the preceding target, leading to its disruption (Breitmeyer, 1984; Breitmeyer & Ganz, 1976). According to feedforward models of masking, there are distinct processing pathways for the target and the mask, each transmitting information at different speeds (Breitmeyer & Ganz, 1976). The mask's faster transmission allows it to interfere with the target's processing, possibly through interchannel inhibitory mechanisms. However, this feedforward explanation appears less plausible when considering the limited neural evidence supporting interchannel inhibition as a mechanism for masking (Enns & Di Lollo, 2000; Fahrenfort et al., 2007). Moreover, the traditional feedforward account was primarily developed to explain metacontrast masking, where the mask does not spatially overlap with the target but still affects its recognition. In contrast, the pattern masks used in our experiments significantly overlapped spatially with the preceding stimuli, suggesting minimal spatial-based inhibition. This is particularly relevant in the higher levels of the ventral stream, where pattern masking is known to impair object recognition (Grill-Spector et al., 2000; Rolls et al., 1999). In these regions, receptive fields cover extensive portions of the visual field (typically around 10-20°; Kobatake and Tanaka, 1994; Rust and DiCarlo, 2010, making spatial-based inhibition almost negligible. Taken together, these considerations significantly challenge the notion of feedforward-based masking as a comprehensive explanation for our findings.

Based solely on literature, a purely feedforward explanation of our observed modulations is not likely. Still, the study paradigms were not specifically designed to completely rule out feedforward interpretations. It is imperative to contemplate alternative methodologies that could more definitively test for recurrent processing in V1, pertinent to both coarse-to-fine processing (Chapter 2) and stimulus degradation (Chapter 4). Investigating separate layers of the cortex could have aided in distinguishing between feedforward and feedback influences in the visual cortex. The segregation of feedforward and feedback inputs across different cortical layers, especially in area V1, offers a unique opportunity for this analysis. Specifically, feedforward connections predominantly target layer 4C of V1, whereas feedback connections are directed primarily towards layers 1-3 and layer 5 (Blasdel & Lund, 1983; Callaway, 1998; Hubel & Wiesel, 1972). While many human neuroimaging techniques are unable to differentiate between these cortical layers, laminar probes, as used in animal tracing and electrophysiology studies, have demonstrated selective targeting of different cortical laminae by neurons from hierarchically distinct areas (Barone et al., 2000; Markov et al., 2014). While highly informative, such techniques are notably invasive. An alternative, non-invasive approach could be ultra-high-field (e.g., 7T) fMRI scanning, which offers sub-millimetre resolution and has potential to discern feedforward and feedback dominated cortical layers. Employing this technique in our experiments would allow for a direct observation of whether certain conditions increase

feedback or feedforward activity. For instance, we could empirically determine whether backward masking specifically disrupts feedback signals, as opposed to inferring this from existing research. This way, the hypothesis that stimulus degradation heightens recurrent processes could be directly investigated by examining activity in the feedback-dominated layers.

Coarse-to-fine processing in behaviour

In the study discussed in chapter 2 (Schuurmans et al., 2023), there is indeed a clear indication of coarse-to-fine processing. Yet, similar trends were notably absent in the behavioural responses of the study documented in chapter 3. This suggests that the translation of the detected coarse-to-fine neural pattern to discernible behavioural responses may not have been as direct as initially anticipated. Various cognitive elements, including attention, anticipation, and task demands, can significantly influence how visual stimuli are processed. These factors might obscure the manifestation of a coarse-to-fine response pattern at a behavioural level. While the detection and categorisation of visual entities might be relatively simultaneous, the identification process requires more time. For example, computer vision algorithms differentiate between face detection and identification (Tsao & Livingstone, 2008) and show it is feasible to acknowledge the presence of a face without discerning its identity. This phenomenon is particularly evident in cases such as prosopagnosia, where individuals can discern faces but fail to recognise them (Barton et al., 2003). This implies a distinction between basic-level detection and deeper, "within-category" identification concerning faces. J. Liu et al. (2013) further support this in an MEG study, and found face-specific responses 100ms post-stimulus introduction. The magnitude of this initial response was linked with successful categorisation, but not recognition, of individual faces. Responses from 170ms onwards correlated with both. Consequently, J. Liu et al. (2013) suggest a two-phase face processing model: initial categorisation followed by identity extraction. Moreover, past evidence indicates that human observers are able to categorise complex and diverse natural scenes above chance with an SOA as short as 12ms with categorisation performance reaching ceiling for SOAs shorter than 80ms (Bacon-Macé et al., 2005). Evidence from past Rapid Serial Visual Presentation (RSVP) studies, in which the perception of each stimulus is masked both by the preceding (forward masking) and following stimulus (backward masking), set a lower bound for the duration necessary to detect, categorise or recognise a face at the individual identity level. Using this approach, Retter et al. (2020) showed that subjects are able to perform face categorisation (i.e., detect the occurrences of face stimuli in a wide diversity of natural images; group-level accuracy at 68%) for presentation durations as short as 50ms. This study, however, did not investigate whether subject could identify the faces. Only a few studies have addressed the timing of face identification and they mainly used RSVP. They suggest that 50ms of face viewing is the minimum stimulus duration to access individual identity cues; identity recognition accuracy is relatively low at such short presentation durations but increases steeply within

the first 100ms of stimulus duration (e.g. Tanskanen et al., 2007). Indeed, our findings from the behavioural study (Chapter 3) corroborate these results; they show that observers are able to recognise the identity of familiar face celebrities above chance even at the shortest SOAs (50ms).

Notably, there is a disparate in the nature of the tasks between the two chapters. The fMRI research (Chapter 2) employed an orthogonal task, concentrating on neural differences in face processing, whereas the behavioural study (Chapter 3) tasked subjects with recognising face identities. The fMRI task was not constructed to reflect the same perceptual demands as the behavioural face recognition task, thereby potentially leading to a discrepancy in the manifestation of coarse-to-fine processing. This fundamental difference in task design introduces a variable that complicates direct comparisons of coarse-to-fine processing across studies. The distinction between mere detection and the deeper process of recognition, along with the differential timing required for these tasks, further muddles the interpretability of the studies' outcomes.

Moreover, the temporal gap between neural activity and behavioural responses adds another layer of complexity. Thorpe et al. (1996) demonstrated that the brain could grasp the essence of a scene within 150ms post-stimulus, as evidenced by event-related potentials. However, behavioural reactions, such as pressing a button, lag considerably behind this neural processing speed. This disparity suggests that the swift neural detection of coarse-to-fine processing may not be directly observable in the slower behavioural responses, which also include the time-consuming motor planning and execution phases. Such timing discrepancies highlight that the findings at the neural level may not seamlessly translate to the behavioural domain, due to the additional processing stages involved in motor responses.

Simplified investigations into coarse-to-fine integration

The methodology employed in this thesis for examining coarse-to-fine integration in visual processing was notably complex. While simpler experimental designs, such as the use of filtered stimuli without backward masking, might suffice for investigating the precedence of coarse-to-fine processing, they fall short in comprehensively testing the integration of SF processing in the brain. An alternative approach could have been to simplify the experimental design by dividing the study into multiple, more focused experiments. This would allow for an isolated examination of specific aspects of the hypothesis, such as the sole impact of backward masking on LSF or HSF processing, instead of merging various complex elements such as temporal dynamics, contrast negation, and analyses across different brain regions. In our research, the initial study (Chapter 2) embodied this complexity by incorporating all these components, whereas the second study (Chapter 3) streamlined the approach, focusing on behavioural data and excluding contrast negation while still involving temporal modulation. However, given that coarse-to-fine integration inherently involves temporal modulations, omitting this aspect would have compromised the depth of our investigation.

The interpretation of the fMRI data from Chapter 2 presented significant challenges. An alternative such as EEG could have offered a simpler means to measure brain activity in response to visual stimuli, potentially providing valuable temporal insights about the processing stages that complement fMRI data. In contrast to a blocked fMRI design, studying event-related potentials (ERPs) via EEG might have simplified data interpretation. Nevertheless, EEG's limited spatial resolution, especially in pinpointing specific activities in regions such as V1, poses a substantial drawback. Given that Petras et al. (2019) already examined coarse-to-fine integration using EEG and later magnetoencephalography (MEG; Petras et al., 2021), replicating their approach would have added little novelty to our findings, nor the implications of coarse-to-fine integration in V1.

An alternative theoretical approach could involve utilising computational models, such as convolutional neural networks, to simulate visual processing in the human brain. These networks can be structured to mimic the hierarchical processing of visual stimuli in the brain, with layers corresponding to different stages of visual analysis, from the initial processing of LSFs to the integration of HSFs. To study the integration of SF processing and its temporal dynamics, a neural network model could be constructed to simulate the experimental conditions of backward masking and filtered stimuli. By adjusting the network's parameters, we could explore how variations in LSF and HSF inputs affect the processing speed and accuracy of the network, providing a theoretical parallel to the human brain's processing. This approach would allow to test key hypotheses about coarse-to-fine integration without the physical and temporal constraints of fMRI or EEG experiments. Moreover, the neural network model could be trained on datasets comprising faces with varying degrees of SF information, similar to the stimuli used in the experiments. By systematically manipulating the input images' SF content and observing the network's output, it is possible to gain insights into how the brain might prioritise different SFs during face recognition tasks. For instance, to assess the model's performance in identifying faces under restricted computations to mimic conditions of temporal masking, akin to the behavioural experiments in Chapter 3. This would enable a direct comparison between the model's predictions and the empirical data, offering a robust test of the coarse-to-fine theory. Furthermore, the neural network's layer-by-layer analysis could offer a closer look at the stages of visual processing, analogous to observing different brain regions in fMRI. While a neural network lacks the spatial resolution of fMRI, its ability to dissect the processing stages offers a unique window into the sequential integration of visual information. The model could potentially show how early layers, akin to V1, handle initial coarse information, and finer details are integrated, mirroring the processing in the human visual system.

Predictive coding framework

The predictions formulated in chapter 2 regarding the sensitivity of the FFA merit further examination, particularly in light of their alignment with the observed responses in V1. In V1, the rationale was clear: lower responses correlate with efficient processing, and disruption of an SF range critical to processing would lead to increased responses. This implies that, since LSFs are important early in time, masking LSFs at early SOAs induce increased responses, while masking at later SOAs would lead to decreased responses. Conversely, for HSFs, which are processed later, the pattern is reversed. The cross-like pattern of SF masking effects that was proposed as the "signature" of coarse-to-fine processing in V1.

The complexity arises when considering higher-level regions such as the FFA, a category-selective region. Initially, it was hypothesised that there would be an overall increase in responses as a function of SOA, with an increase that was expected to be more pronounced for HSF masking, creating the cross-like pattern. However, this prediction might not be straightforward. One might reason that disrupting a crucial SF range should decrease responses in category-selective regions, as the disruption would make the stimulus less representative of the category it is tuned to. Contrary to these assumptions, we observed the cross-like pattern with a steeper increase for HSF masking, a point not extensively discussed in our published work. The observed results might be interpreted through the lens of predictive coding, though it is important to note that such an interpretation is speculative and represents a post hoc interpretation of the data.

According to the predictive coding framework, the brain functions by anticipating sensory information based on past experiences, thereby enabling efficient processing of predicted (or expected) information and allocating resources to novel or unexpected stimuli (Friston, 2005; Mumford, 1992; Rao & Ballard, 1999). This framework posits that the brain constructs models of the world using context and memory to predict sensory input (e.g. Figure 5.3). Higher-level cortical areas develop a predictive model, which is conveyed to lower sensory areas through feedback connections. Meanwhile, feedforward connections convey the error signal - the discrepancy or "mis-match" between anticipated information and actual sensory input (Rao & Ballard, 1999), allowing for constant model refinement based on this feedback. In support of the predictive coding framework, Alink et al. (2010) conducted a study on the primary visual cortex, where they found that predictable stimuli generated a lower BOLD signal compared to unpredictable ones. They used an apparent motion illusion, presenting an additional bar along the path of the illusory moving bar at either predictable or unpredictable times. The predictable stimulus led to a lower BOLD response, aligning with the concept of predictive coding, where lower level predictions are suppressed by higher-level feedback. Another study by den Ouden et al. (2010) provides further insights into the brain's sensitivity to short-term contingencies. This study showed that when an auditory tone consistently preceded either a face or a house, brain activation in the areas that process these images (the FFA for faces and the PPA for houses) decreased. This reduction occurred even

though the association between the tone and image was artificially created and the probability of their pairing changed throughout the experiment. Subjects were unaware of when these changes would happen, yet a clear negative correlation was observed between neural response and stimulus predictability. When considering the findings of both Alink et al. (2010) and den Ouden et al. (2010), it becomes evident that the brain's response to predictability varies across time scales, experimental contexts, and visual hierarchy levels. This aligns with the hypothesis by Rao and Ballard (1999), stating that higher brain areas feed predictions back to compare with incoming sensory signals, this leads lower areas to represent the discrepancy or "error signal" rather than the stimulus itself.



FIGURE 5.3: Monochrome image. Viewing this high-contrast monochrome image may not allow for immediate recognition of its content. Within the predictive coding framework, the deficit of colour and nuanced detail hinders the brain's capacity to generate an accurate predictive model. Exposure to the coloured counterpart of this image (see Figure 5.4) furnishes the necessary contextual information, enabling the brain to construct the appropriate predictions for successful identification.

Murray et al. (2002) highlighted that reduced brain activity might signify either a general decrease in the strength of a representation or a more efficient encoding of predictable stimuli, marked by lower average activation but a higher SNR. In essence, the predictions of specific features or object categories can heighten fMRI BOLD activity in the brain regions typically involved in processing those features or categories (Esterman & Yantis, 2010; Kok et al., 2014; Puri et al., 2009). Additionally, the (feedforward) error signals may elevate fMRI BOLD activity in instances of a mismatch between bottom-up sensory input and top-down predictive signals.

Revisiting the hypothesis pertaining to the FFA within the predictive coding framework, the responses observed can be interpreted as a combination of signal and prediction error. Considering the predictive signal in the FFA, long processing time for an intact face results in a robust prediction, leading to heightened responses. On the other hand, in terms of activity attributable to prediction error; when a mask coincides with a critical processing period for a specific SF range, a bigger mis-match and thus a greater error are anticipated. In terms of predictions this means that, masking LSFs at early SOAs results in a weak prediction accompanied by a significant error. And at later SOAs, a strong prediction aligns with negligible error. In other words, an increased slope as a function of SOA for LSF masking responses. On the other hand, HSF masking at short SOAs is characterised by both a weak prediction and a minimal error, while late masking stages would involve a strong prediction coupled with a substantial error. This pattern is expected to yield a steeper increasing response curve, aligning with our findings.

However, the predictive coding theory presents complexities. As previously discussed, high activity in a region, such as the fMRI BOLD signal, could signify an error signal. This heightened signal might also represent a predictive signal, particularly in category-selective regions. This dual interpretation within the predictive coding framework poses a challenge, as it allows for various findings to be accommodatively explained by the theory. Consequently, this adaptability makes it difficult to refute the theory, which presents challenges in testing hypotheses within this framework.

Statistical Linear Mixed effects Modelling approach

The application of mixed-effects models for experimental data analysis has gained significant traction, as outlined by Baayen et al. (2008) who discuss this method using the lme4 package in R (Bates et al., 2014). Researchers transitioning from traditional analysis methods such as ANOVAs to mixed models often encounter challenges due to the less intuitive aspects of mixed modelling. A particular hurdle is determining the significance of fixed effects in the model output, especially with complex designs such as crossed or unbalanced data sets. Baayen et al. (2008) highlight the difficulties in defining the denominator degrees of freedom (DenDF) for models with multiple levels. The conundrum lies in whether to consider the number of observations, subjects, items, grouping factors, or a combination for DenDF calculation.

For instance, in chapter 2 of this thesis, we observed DenDF exceeding 10,000 in both V1 and FFA regions, reflecting the complexity inherent in these calculations. Unlike simpler ANOVA models, DenDF in mixed models often employs sophisticated methods such as Satterthwaite's approximation (Hrongs-Tai Fai & Cornelius, 1996; Satterthwaite, 1941) or Kenward-Roger's approximation (Halekoh & Højsgaard, 2014; Kenward & Roger, 1997). These methods account for the random effects structure in estimating the effective sample size for each fixed effect test, which can lead to significantly large DenDFs. The advantage of

Satterthwaite's approach is its adjustment for data complexity, thereby negating the need for further corrections. The substantial DenDF in our model implies that our error estimates are based on a robust data set, enhancing the reliability and precision of the F-statistics and p-values. This lends greater credibility to the statistical significance of our findings, affirming that observed effects are genuine and not mere by-products of insufficient statistical power. In traditional statistical tests, an excessively large DenDF relative to the sample size might raise concerns of overfitting or similar issues. However, in mixed models, particularly those involving numerous subjects and repeated measures (or voxels in our case), such large DenDFs are both expected and appropriate, signifying the model's capacity to utilise all available data effectively rather than overfitting.

On the other hand, there were particular small Numerator Degrees of Freedom (NumDF) in the results of chapter 2. NumDF reflects the specific and focused nature of the hypotheses tested. A NumDF of 1 for the interaction terms means that each interaction term was examined as a singular hypothesis about the joint effect of these factors. This simplifying our interpretation to whether the effect of one variable significantly varies at different levels of another.



FIGURE 5.4: *Colourised version of pervious photograph. Upon observing the colourised counterpart of the previous image (Figure 5.3), it becomes apparent the photograph depicts a dog laying amidst a hedge. The colours and details furnish the brain with critical contextual information, facilitating the construction of an accurate predictive model. This in turn enables the recognition of the dog in the high-contrast monochrome depiction of Figure 5.3.*

Timing in the visual system

Much of our knowledge about the neural temporal dynamics of vision are derived from single-unit studies in macaque monkeys. These animals, whose visual systems functionally resemble those of humans, have been studied more extensively (Bullier, 2001; Schmolesky et al., 1998). These studies often employ rate coding, suggesting that neurons convey information through alterations in their firing rate. However, other aspects, such as spike timing, can also convey essential data (Rieke et al., 1996). When presented with a static visual stimulus, there is an initial delay before the number of spikes increases. This delay, known as latency, is short in V1 or V2 and lengthens in higher visual regions such as the IT and frontal cortex (Figure 5.5; Bullier, 2001; Lamme et al., 2002; Schmolesky et al., 1998). By approximately 190ms post-stimulus onset, most visually responsive neurons in the macaques are activated. Human brain latencies are generally longer, which is believed to be due to the larger physical size of the human brain and longer distances that neuronal signals must cover (Yoshor et al., 2007). Response latencies in specific areas can vary greatly and are affected by various factors, making them challenging to compare across studies. The firing rate of a visual cortical neuron, after detecting a static stimulus, rises sharply and then diminishes, stabilising at a reduced level. The initial increase is attributed to feed-forward inputs, while the post-rise phase is shaped majorly by recurrent processing (Lamme & Roelfsema, 2000; Scholte et al., 2006). Such research on latencies enriches our understanding of the brain's processing duration to formulate a visual perception.

While the foundational studies delved deeply into neural latencies, the primary emphasis of this thesis shifts to the processing duration, which delineates the span that the visual system allocates to interpret a visual stimulus. Our methodology incorporated backward masking, a technique that facilitates distinguishing minor variances in recurrent processes as a function of stimulus exposure duration. The findings elaborated upon in chapters 2 and 3 show modulation of responses due to backward masking as a function of SOA ranging between 50 and 150ms. The temporal span is based on the research by Goffaux et al. (2011), where they investigated coarse-to-fine processing using backward masking, and using SF filtered faces. It was observed that face-preferring regions responded strongly to coarse/LSF face information after about 75ms of processing. As SOA increased, LSF processing decreased, mostly until 150ms, while the processing of fine HSF face information became more substantial over time. Indeed, more studies showed as well that 150ms of processing is enough to categorise visual stimuli (e.g., Bachmann and Allik, 1976; Grill-Spector and Kanwisher, 2005; Thorpe et al., 1996). For example, Grill-Spector and Kanwisher (2005) examined the time required to perform different types of object recognition tasks by measuring performance as a function of stimulus duration. In their study, object detection and categorisation tasks required a similar amount of processing time (~50ms), indicating that categorisation of the object occurs before identification. For a stimulus duration of 167ms, the performance in all tasks, including identification, was close to perfect. However, a study by Bachmann et al. (2005), showed

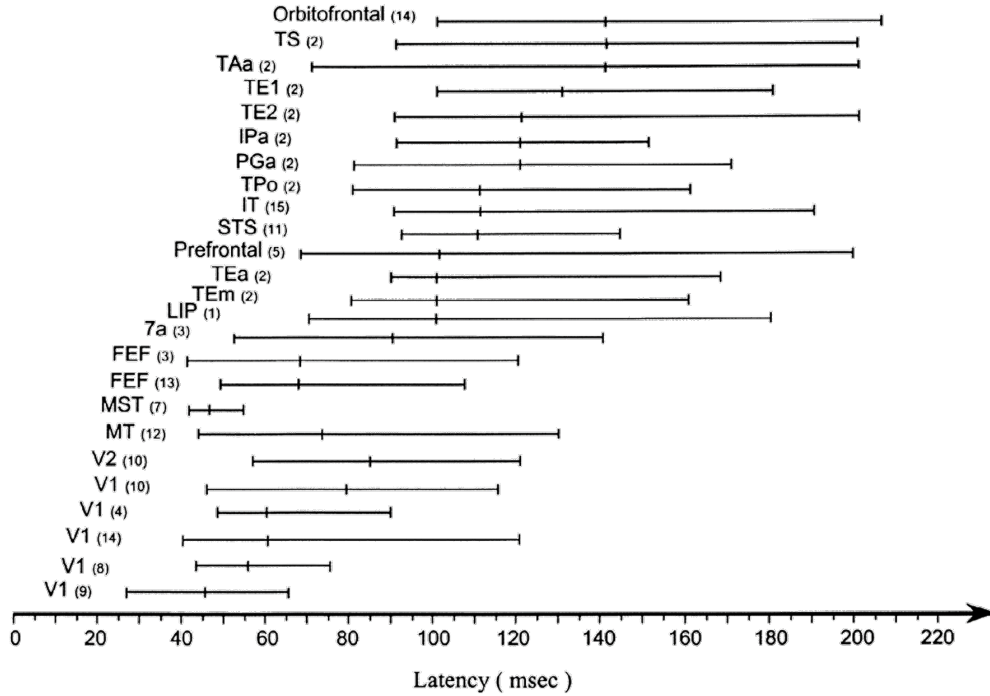


FIGURE 5.5: Visual response latencies across various cortical areas based on monkey physiology; adapted from Bullier (2001). The middle tick indicates the median latency, while the outer ticks represent the 10th and 90th percentiles. References for the data (between brackets) can be found in Bullier (2003). It is noteworthy that the briefest latencies are not exclusively linked to the initial hierarchy stages. Specifically, areas such as MT, MST, and FEF exhibit minimal latencies, even though they are positioned at upper tiers in the hierarchy's early model.

longer processing duration for identification. In this study, they briefly presented two geometric shapes to subjects. When displayed simultaneously (0ms SOA), they appeared as one blended shape. Identification of the initial target shape was hardest with SOAs of 40-80ms, hinting at the importance of recurrent processing during this period for identification. A clear, distinct perception of the target required an SOA of 150ms, and accurate identification took about 250ms, marking the completion of the corresponding processes. This finding implies that, while an SOA of 150ms might suffice for basic processing in some contexts, visual processing continues beyond this mark. Indeed, beyond the initial 150ms, Hansen et al. (2021) noted a more intricate pattern of SF sensitivity beyond, involving alternating periods of coarse-to-fine and fine-to-coarse processing. They utilised EEG's high temporal resolution to create an encoding technique based on the responses generated by a population to real-world scenes. Their analyses revealed a series of non-uniform transformations prioritising different SFs at different locations of scenes over time, suggesting a more complex temporal coding of visual information than a straightforward coarse-to-fine analysis, beyond 150ms of processing. Given the detailed research on the temporal dynamics of vision in both

macaque monkeys and humans, and the clear evidence pointing to significant visual processing occurring within a span of 150ms, the choice to utilise SOAs ranging between 50 and 150ms in our experiments is both informed and grounded in this broader understanding of neural latencies and visual processing durations.

Face stimuli in research

Research on recurrent or coarse-to-fine processing has often used basic stimuli such as dots, gratings, or simple shapes (e.g., Breitmeyer, 1975; Hughes et al., 1996; Jones and Keck, 1978; Kok and de Lange, 2014; Lamme, 1995; T. S. Lee and Nguyen, 2001; Mihaylova et al., 1999; Musselwhite and Jeffreys, 1985; Parker, 1980; Parker and Dutch, 1987; Watt, 1987; Zipser et al., 1996). Yet, in naturalistic vision, inputs are a blend of different SFs. In our studies, we sought to employ more ecologically relevant stimuli. To robustly engage the entire ventral visual pathway, we chose face stimuli, a well-studied visual category (e.g., Fox et al., 2009; Gauthier et al., 2000; Grill-Spector et al., 2017, 2018; Ishai, 2008). The faces we used in our experiments were carefully controlled: luminance and contrast balanced, grey-scaled, and presented within an oval frame. Notably, the facial images in chapters 2 and 4 excluded features such as ears and hair. In other words, although these face stimuli were more natural than simple dots or gratings, they were not entirely naturalistic.

Furthermore, the stimuli were embedded in an artificial (phase-scrambled) background, detracting them further from naturalistic settings (i.e., faces with a meaningful background). An alternative might involve placing the face or an object within a natural scene. Groen et al. (2018) found that when stimuli are embedded in high-complexity scenes during an animal detection task, there was more engagement of feedback as compared to simpler scenes. In other words, enhanced recurrent computations were needed for object recognition (Groen et al., 2018; Rajaei et al., 2019). While naturalistic scenes hold many merits with regards to studying recurrent processing (Felsen & Dan, 2005; Felsen et al., 2005; Talebi & Baker, 2012), they present challenges in maintaining control due to the numerous varying (and potentially unknown) dimensions. For instance, spectral properties might align with other contextual elements within the scene, as spatial coherence has been shown to correlate with the perceived naturalness of a scene (Rajaei et al., 2019) and could provide diagnostic information for object detection. Despite its limitations, embedding a stimulus in a scene could be a worthwhile consideration when aiming for more ecologically valid stimuli compared to those combined with artificial noise.

To further enhance the realism of faces, we must consider real life scenarios; faces move. Incorporating facial motion into studies offers profound insights into face perception. Beyond merely aiding facial expression recognition, facial motion has been shown to bolster speech comprehension when auditory senses are hindered (L. E. Bernstein et al., 2000; Rosenblum et al., 2002). Motion is also instrumental in discerning aspects such as gender and identity (e.g., Hill and Johnston, 2001; O'Toole et al., 2002). Interestingly, the richness of identity

cues drawn from facial movements varies with the type of movement, with conversational actions offering more distinctive identity expressions than emotional ones (Dobs et al., 2016). Processing moving faces is believed to engage a broad spectrum of visual areas ranging from the middle occipital and temporal gyri and progressing through the bilateral superior temporal sulcus (Allison et al., 2000) to frontal areas such as the inferior and middle frontal gyri (M. Bernstein and Yovel, 2015; Foley et al., 2012). Furthermore, Mermillod et al. (2019) utilised a neural network model to emphasise the role of recurrent connections in predicting dynamic emotions, suggesting that top-down connectivity from associative areas improves the anticipation of dynamic emotional expressions. Grasping these dynamics in facial recognition and their neural foundations will undoubtedly deepen our comprehension of human visual processing in real-world scenarios.

The studies presented in this thesis raise a question in relation to motion in combination with SF processing, a pivotal element in our research. LSFs are primarily processed through the magnocellular pathway (for more information see chapter 1), which is also responsible for processing motion information (Livingstone & Hubel, 1988). This leads to an important consideration: the potential impact of using static images on the temporal properties of the stimuli and the resultant effects on recurrent processing. Specifically, it prompts an evaluation of whether the exclusive use of static images in our experiments might limit our ability to fully explore all facets of recurrent processing. The relevance of this consideration is underscored by research indicating that face recognition is more efficient and accurate when dynamic faces are presented during the learning phase, followed by static faces during testing (Butcher et al., 2011; M. Zhao et al., 2014). Additionally, ERP studies have shown that the P1 amplitude (positive ERP component at after 100ms post-stimulus onset associated with face processing) is higher for LSF faces compared to HSF faces, with this effect being more pronounced for moving faces than for static ones (H. Wang et al., 2023). Furthermore, while neuroimaging studies have found no difference in the FFA's response to moving versus static faces, the dorsal pathway, extending from the middle temporal area to the superior temporal sulcus, showed greater activation in response to moving faces (M. Bernstein & Yovel, 2015; Pitcher et al., 2011). These findings collectively suggest that the dynamic aspects of visual stimuli play a role in processing and recognition. Consequently, the use of static images in our studies might limit the exploration of the full spectrum of recurrent processing, particularly in aspects that involve motion and temporal dynamics. This raises the possibility that additional studies incorporating dynamic stimuli might provide a more comprehensive understanding of recurrent processing in the visual system.

Beyond faces

In this thesis, we employed face stimuli for their established robust interaction with a significant portion of the visual system and the resulting pronounced recurrent processes (Fox et al., 2009; Gauthier et al., 2000; Grill-Spector et al., 2017, 2018; Ishai, 2008). Evidence from both behavioural studies and fMRI research suggests that face processing is particularly dependent on SFs compared to the processing of other object categories (Collin et al., 2004; Yue et al., 2006). For instance, Davies et al. (1978) observed that it is more difficult to recognise faces from line drawings as compared to photographs. In addition, an even earlier study Ryan and Schwartz (1956) suggests object recognition is easier in line drawings over photographs. Bruce et al. (1992) and Hayes (1988) found reduced recognition rate for faces in line drawings and two-tone images. This contrast is often attributed to the reliance of object recognition on defined features such as edges and lines Biederman and Kalocsais (1997), as opposed to facial recognition, which necessitates a richer array of visual cues including edges, textures, and shading (Bruce et al., 1992).

Notably, it is proposed that, particularly for faces, SF-dependent representations originating in V1 remain distinct during advanced processing stages (Biederman & Kalocsais, 1997). Faces characteristically exhibit more energy in low than in high SFs, a pattern more distinct than in other visual categories. This may be the reason why essential behavioural and neural indicators of specialised face processing predominantly rely on the processing of LSF information (Goffaux et al., 2003; Goffaux & Rossion, 2006). Such LSF predominance in faces potentially facilitates the coarse-to-fine processing mechanism. However, mechanisms such as recurrent processing are not exclusive to faces, having been studied in the categorisation of degraded stimuli, such as distinguishing animals from other objects (O'Reilly et al., 2013; Wyatte et al., 2012). The concept of coarse-to-fine processing has also been investigated in scene processing contexts (Kauffmann, Chauvin, Pichat, & Peyrin, 2015; Kauffmann, Ramanoël, et al., 2015). Our 3T fMRI study (Chapter 2) further supports this notion, showing the involvement of non-face selective regions such as V1, dorsal, and orbitofrontal areas in coarse-to-fine processing. This resonates with the findings that object recognition hinges on the dorsal and frontal areas (Ayzenberg & Behrmann, 2022; Bar, 2004; Kay & Yeatman, 2016; Zachariou et al., 2017), which rapidly produce a coarse, initial impression of incoming visuals (Ayzenberg & Behrmann, 2022; Ayzenberg et al., 2023; Bar et al., 2006; Fintzi & Mahon, 2014; Kauffmann, Bourgin, et al., 2015; Kauffmann, Chauvin, Pichat, & Peyrin, 2015; Kveraga et al., 2007; Peyrin et al., 2010; Zachariou et al., 2017). In essence, the outcomes from our studies likely extend beyond face stimuli, suggesting broader research applicability.

Concluding remarks

Cognitive neuroscience aims to unravel the complex mechanics underlying human cognition and its physiological base. A significant aspect of this field is the examination of feedforward signals along with other fundamental brain mechanisms such as feedback and lateral signals (i.e. recurrent processing). The use of backward masking with complex visual stimuli has been instrumental in facilitating the precise temporal examination of recurrent processes. When integrated with fMRI, it enabled a detailed examination of these processes in specific brain regions. The focus of this thesis is on complex, naturalistic stimuli, specifically how the brain processes its basic elements, such as SFs, which are proposed to be handled in a coarse-to-fine manner. Unveiling that not only V1 but also various brain regions (ventral, dorsal, and frontal) operate in this refining manner with full-spectrum face stimuli, this study contributes to the understanding of visual processing in the brain. Moreover, this thesis additionally shows the increased engagement of recurrent processing when coping with noisy visual stimuli, further underscoring the dynamic and adaptable nature of the visual system. While these insights advance our understanding, they also pose new questions such as "Will coarse-to-fine masking have a similar effect on other stimuli as it does on faces?" and "How do various brain regions interact beyond 150ms of stimulus processing?". This thesis offers a step towards deciphering the dynamic nature of visual perception and contributes to our understanding of the profound complexity of the human brain.

Publications and presentations

Articles published in peer-reviewed journals

Schuermans, J. P., Bennett, M. A., Petras, K., & Goffaux, V. (2023). Backward masking reveals coarse-to-fine dynamics in human V1. *NeuroImage*, 274, 120139.

Conferences with a scientific selection committee

Oral

Schuermans, J. P., Bennett, M. A., Dey, M., & Goffaux, V. (2023). Backward masking coarse-to-fine processing of faces in the human visual system. European Conference on Visual Perception, Paphos, Cyprus.

Poster

Dey, M., **Schuermans, J.P.**, & Goffaux, V. (2023, November 23-24). Backward masking coarse-to-fine processing of faces in the human visual system. Neurocog, Brussels, Belgium.

Canoluk, M.U., **Schuermans, J.P.**, Bennett, M.A., Grandjean, M., Moors, P., & Goffaux, V. (2023, August 27-31). Contributions of low-level mechanisms to (non)face-specialized contextual mechanisms. European Conference on Visual Perception, Paphos, Cyprus.

Dey, M., **Schuermans, J.P.**, & Goffaux, V. (2023, August 27-31). Backward masking coarse-to-fine processing of faces in the human visual system. European Conference on Visual Perception, Paphos, Cyprus.

Schuermans, J., Bennett, M., & Goffaux, V. (2022). Coarse-to-fine processing of faces in the core face network and V1. *Journal of Vision*, 22(14), 4045-4045.

Schuermans, J. P., Bennett, M., & Goffaux, V. (2022, April 28-13). Coarse-to-fine processing of faces in V1 [conference poster]. NVP winter conference, Egmond aan Zee, Netherlands.

Schuermans, J. P., Bennett, M., & Goffaux, V. (2021, November 23-24). Coarse-to-fine processing of faces throughout the ventral visual hierarchy. Neurocog, Louvain-la-Neuve, Belgium.

Schuermans, J. P., Bennett, M., & Goffaux, V. (2021). Coarse-to-fine processing of faces throughout the ventral visual hierarchy. *Journal of Vision*, 21(9), 2651-2651.

Schuermans, J. P., Scholts, T., & Goffaux, V. (2019, October 10-11). Face Processing in V1: Coarse-to-Fine? GDR Vision, Marseille, France.

Schuermans, J. P., Scholts, T., & Goffaux, V. (2019). Face Processing in V1: Coarse-to-Fine? *Perception*, 48(S2), 133-133.

References

- Aguirre, G. K., Zarahn, E., & DEsposito, M. (1998). An area within human ventral cortex sensitive to building stimuli: Evidence and implications. *Neuron*, *21*(2), 373–383.
- Ahissar, M., & Hochstein, S. (2004). The eureka effect. *Trends in Cognitive Sciences*, *10*(8), 457–464.
- Ahmed, B., Hanazawa, A., Undeman, C., Eriksson, D., Valentiniene, S., & Roland, P. E. (2008). Cortical dynamics subserving visual apparent motion. *Cereb. Cortex*, *18*(12), 2796–2810.
- Akrami, A., Liu, Y., Treves, A., & Jagadeesh, B. (2009). Converging neuronal activity in inferior temporal cortex during the classification of morphed stimuli. *Cerebral Cortex*, *19*(4), 760–776.
- Albrecht, D. G., Geisler, W. S., Frazor, R. A., & Crane, A. M. (2002). Visual cortex neurons of monkeys and cats: Temporal dynamics of the contrast response function. *Journal of neurophysiology*, *88*(2), 888–913.
- Ales, J. M., Farzin, F., Rossion, B., & Norcia, A. M. (2012). An objective method for measuring face detection thresholds using the sweep steady-state visual evoked response. *Journal of vision*, *12*(10), 18–18.
- Alink, A., Schwiedrzik, C. M., Kohler, A., Singer, W., & Muckli, L. (2010). Stimulus predictability reduces responses in primary visual cortex. *Journal of Neuroscience*, *30*(8), 2960–2966.
- Allen, E. A., & Freeman, R. D. (2006). Dynamic spatial processing originates in early visual pathways. *Journal of Neuroscience*, *26*(45), 11763–11774.
- Allison, T., Puce, A., & McCarthy, G. (2000). Social perception from visual cues: Role of the sts region. *Trends in cognitive sciences*, *4*(7), 267–278.
- Anderson, J. C., & Martin, K. A. (2009). The synaptic connections between cortical areas v1 and v2 in macaque monkey. *Journal of Neuroscience*, *29*(36), 11283–11293.
- Andersson, J. L., Skare, S., & Ashburner, J. (2003). How to correct susceptibility distortions in spin-echo echo-planar images: Application to diffusion tensor imaging. *Neuroimage*, *20*(2), 870–888.
- Andrews, T. J., & Ewbank, M. P. (2004). Distinct representations for facial identity and changeable aspects of faces in the human temporal lobe. *Neuroimage*, *23*(3), 905–913.
- Angelucci, A., Bijanzadeh, M., Nurminen, L., Federer, F., Merlin, S., & Bressloff, P. C. (2017). Circuits and mechanisms for surround modulation in visual cortex. *Annual review of neuroscience*, *40*, 425–451.
- Angelucci, A., & Bressloff, P. C. (2006). Contribution of feedforward, lateral and feedback connections to the classical receptive field center and extra-classical receptive field surround of primate v1 neurons. *Progress in brain research*, *154*, 93–120.

- Arcurio, L. R., Gold, J. M., & James, T. W. (2012). The response of face-selective cortex with single face parts and part combinations. *Neuropsychologia*, 50(10), 2454–2459.
- Astorga, G., Chen, M., Yan, Y., Altavini, T. S., Jiang, C. S., Li, W., & Gilbert, C. (2022). Adaptive processing and perceptual learning in visual cortical areas V1 and V4. *Proc. Natl. Acad. Sci. U. S. A.*, 119(42), e2213080119.
- Avidan, G., Tanzer, M., Hadj-Bouziane, F., Liu, N., Ungerleider, L. G., & Behrmann, M. (2014). Selective dissociation between core and extended regions of the face processing network in congenital prosopagnosia. *Cerebral cortex*, 24(6), 1565–1578.
- Ayzenberg, V., & Behrmann, M. (2022). Does the brain’s ventral visual pathway compute object shape? trends in cognitive sciences. *Trends in Cognitive Sciences*.
- Ayzenberg, V., Simmons, C., & Behrmann, M. (2023). Temporal asymmetries and interactions between dorsal and ventral visual pathways during object recognition. *Cerebral Cortex Communications*, 4(1), tgad003.
- Ayzenshtat, I., Gilad, A., Zurawel, G., & Slovin, H. (2012). Population response to natural images in the primary visual cortex encodes local stimulus attributes and perceptual processing. *Journal of Neuroscience*, 32(40), 13971–13986.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language*, 59(4), 390–412.
- Bachmann, T. (1991). Identification of spatially quantised tachistoscopic images of faces: How many pixels does it take to carry identity? *European Journal of Cognitive Psychology*, 3(1), 87–103.
- Bachmann, T., & Allik, J. (1976). Integration and interruption in the masking of form by form. *Perception*, 5(1), 79–97.
- Bachmann, T., Luiga, I., & Pöder, E. (2005). Variations in backward masking with different masking stimuli: Ii. the effects of spatially quantised masks in the light of local contour interaction, interchannel inhibition, perceptual retouch, and substitution theories. *Perception*, 34(2), 139–154.
- Bacon-Macé, N., Macé, M. J.-M., Fabre-Thorpe, M., & Thorpe, S. J. (2005). The time course of visual processing: Backward masking and natural scene categorisation. *Vision research*, 45(11), 1459–1469.
- Bair, W., Cavanaugh, J. R., & Movshon, J. A. (2003). Time course and time-distance relationships for surround suppression in macaque v1 neurons. *Journal of Neuroscience*, 23(20), 7690–7701.
- Bankó, É. M., Gál, V., Körtvélyes, J., Kovács, G., & Vidnyánszky, Z. (2011). Dissociating the effect of noise on sensory processing and overall decision difficulty. *Journal of Neuroscience*, 31(7), 2663–2674.
- Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., Hämäläinen, M. S., K, M., Schacter, D. L., Rosen, B. R., & Halgren, E. (2006). Top-down facilitation of visual recognition. *Proc. Natl. Acad. Sci. U. S. A.*, 103(2), 449–454.

- Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience*, 5(8), 617–629.
- Barlow, H. B. (1975). Visual experience and cortical development. *Nature*, 258(5532), 199–204.
- Barlow, H. (2001). Redundancy reduction revisited. *Network: computation in neural systems*, 12(3), 241.
- 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 Neuroscience*, 20(9), 3263–3281.
- Barragan-Jason, G., Cauchoix, M., & Barbeau, E. (2015). The neural speed of familiar face recognition. *Neuropsychologia*, 75, 390–401.
- Bartels, A. (2014). Visual perception: Early visual cortex fills in the gaps. *Current Biology*, 24(13), R600–R602.
- Barton, J. J., Cherkasova, M. V., Press, D. Z., Intriligator, J. M., & OConnor, M. (2003). Developmental prosopagnosia: A study of three patients. *Brain and cognition*, 51(1), 12–30.
- Baseler, H., & Sutter, E. (1997). M and p components of the vep and their visual field distribution. *Vision research*, 37(6), 675–690.
- Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron*, 76(4), 695–711.
- Bastos, A. M., Vezoli, J., Bosman, C. A., Schoffelen, J.-M., Oostenveld, R., Dowdall, J. R., De Weerd, P., Kennedy, H., & Fries, P. (2015). Visual areas exert feedforward and feedback influences through distinct frequency channels. *Neuron*, 85(2), 390–401.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.
- 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.
- Benevento, L., Creutzfeldt, O., & Kuhnt, U. (1972). Significance of intracortical inhibition in the visual cortex. *Nature New Biology*, 238(82), 124–126.
- Benton, A. L. (1980). The neuropsychology of facial recognition. *American Psychologist*, 35(2), 176.
- Ben-Yishai, R., Bar-Or, R. L., & Sompolinsky, H. (1995). Theory of orientation tuning in visual cortex. *Proceedings of the National Academy of Sciences*, 92(9), 3844–3848.
- Bernstein, L. E., Tucker, P. E., & Demorest, M. E. (2000). Speech perception without hearing. *Perception & Psychophysics*, 62(2), 233–252.
- Bernstein, M., Erez, Y., Blank, I., & Yovel, G. (2018). An integrated neural framework for dynamic and static face processing. *Sci. Rep.*, 8(1), 7036.
- Bernstein, M., & Yovel, G. (2015). Two neural pathways of face processing: A critical evaluation of current models. *Neuroscience & Biobehavioral Reviews*, 55, 536–546.

- Bex, P. J., & Makous, W. (2002). Spatial frequency, phase, and the contrast of natural images. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.*, 19(6), 1096–1106.
- Biederman, I., & Cooper, E. E. (1992). Size invariance in visual object priming. *Journal of Experimental Psychology: Human Perception and Performance*, 18(1), 121.
- Biederman, I., & Kalocsais, P. (1997). Neurocomputational bases of object and face recognition. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 352(1358), 1203–1219.
- Bindemann, M., Attard, J., Leach, A., & Johnston, R. A. (2013). The effect of image pixelation on unfamiliar-face matching. *Applied Cognitive Psychology*, 27(6), 707–717.
- Blasdel, G. G., & Lund, J. S. (1983). Termination of afferent axons in macaque striate cortex. *Journal of Neuroscience*, 3(7), 1389–1413.
- Boehler, C., Schoenfeld, M., Heinze, H.-J., & Hopf, J.-M. (2008). Rapid recurrent processing gates awareness in primary visual cortex. *Proceedings of the National Academy of Sciences*, 105(25), 8742–8747.
- Boutet, I., Collin, C., & Faubert, J. (2003). Configural face encoding and spatial frequency information. *Perception & Psychophysics*, 65, 1078–1093.
- Boynton, G. M., Engel, S. A., Glover, G. H., & Heeger, D. J. (1996). Linear systems analysis of functional magnetic resonance imaging in human v1. *Journal of Neuroscience*, 16(13), 4207–4221.
- Bredfeldt, C. E., & Ringach, D. (2002). Dynamics of spatial frequency tuning in macaque v1. *Journal of Neuroscience*, 22(5), 1976–1984.
- Breitmeyer, B. G. (1975). Simple reaction time as a measure of the temporal response properties of transient and sustained channels. *Vision Res.*, 15(12), 1411–1412.
- Breitmeyer, B. G. (1984). *Visual masking: An integrative approach*. Oxford University Press.
- Breitmeyer, B. G. (1973). A relationship between the detection of size, rate, orientation and direction in the human visual system. *Vision Research*, 13(1), 41–58.
- Breitmeyer, B. G., & Ganz, L. (1976). Implications of sustained and transient channels for theories of visual pattern masking, saccadic suppression, and information processing. *Psychological review*, 83(1), 1.
- Bridgeman, B. (1980). Temporal response characteristics of cells in monkey striate cortex measured with metacontrast masking and brightness discrimination. *Brain research*, 196(2), 347–364.
- Brigham, E. O. (1988). *The fast fourier transform and its applications*. Prentice-Hall, Inc.
- Bruce, V., Hanna, E., Dench, N., Healey, P., & Burton, M. (1992). The importance of mass in line drawings of faces. *Applied cognitive psychology*, 6(7), 619–628.
- Bruce, V., & Young, A. (1986). Understanding face recognition. *British journal of psychology*, 77(3), 305–327.

- Budd, J. M. (1998). Extrastriate feedback to primary visual cortex in primates: A quantitative analysis of connectivity. *Proc. Biol. Sci.*, 265(1400), 1037–1044.
- Buffalo, E. A., Fries, P., Landman, R., Liang, H., & Desimone, R. (2010). A backward progression of attentional effects in the ventral stream. *Proceedings of the National Academy of Sciences*, 107(1), 361–365.
- Bullier, J., Hupé, J., James, A., & Girard, P. (1996). Functional interactions between areas v1 and v2 in the monkey. *Journal of Physiology-Paris*, 90(3-4), 217–220.
- Bullier, J., McCourt, M., & Henry, G. (1988). Physiological studies on the feedback connection to the striate cortex from cortical areas 18 and 19 of the cat. *Experimental Brain Research*, 70, 90–98.
- Bullier, J. (2001). Integrated model of visual processing. *Brain research reviews*, 36(2-3), 96–107.
- Bullier, J. (2003). Cortical connections and functional interactions between visual cortical areas. *Neuropsychol. Vis*, 251–256.
- 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.
- Burton, M. (2013). Why has research in face recognition progressed so slowly? the importance of variability. *Quarterly Journal of Experimental Psychology*, 66(8), 1467–1485.
- Bushnell, M. C., Goldberg, M. E., & Robinson, D. L. (1981). Behavioral enhancement of visual responses in monkey cerebral cortex. i. modulation in posterior parietal cortex related to selective visual attention. *Journal of Neurophysiology*, 46(4), 755–772.
- Butcher, N., Lander, K., Fang, H., & Costen, N. (2011). The effect of motion at encoding and retrieval for same-and other-race face recognition. *British Journal of Psychology*, 102(4), 931–942.
- Butterworth, S. (1930). On the theory of filter amplifiers. *Wireless Engineer*, 7(6), 536–541.
- Caharel, S., Ramon, M., & Rossion, B. (2014). Face familiarity decisions take 200 msec in the human brain: Electrophysiological evidence from a go/no-go speeded task. *Journal of Cognitive Neuroscience*, 26(1), 81–95.
- Callaway, E. M. (1998). Local circuits in primary visual cortex of the macaque monkey. *Annual review of neuroscience*, 21(1), 47–74.
- Camprodon, J. A., Zohary, E., Brodbeck, V., & Pascual-Leone, A. (2010). Two phases of v1 activity for visual recognition of natural images. *Journal of cognitive neuroscience*, 22(6), 1262–1269.
- Carandini, M., & Ringach, D. L. (1997). Predictions of a recurrent model of orientation selectivity. *Vision research*, 37(21), 3061–3071.
- Chadwick, A., Khan, A. G., Poort, J., Blot, A., Hofer, S. B., Mrsic-Flogel, T. D., & Sahani, M. (2023). Learning shapes cortical dynamics to enhance integration of relevant sensory input. *Neuron*, 111(1), 106–120.

- Cisler, J. M., Bush, K., & Steele, J. S. (2014). A comparison of statistical methods for detecting context-modulated functional connectivity in fmri. *Neuroimage*, *84*, 1042–1052.
- Cochran, W. T., Cooley, J. W., Favin, D. L., Helms, H. D., Kaenel, R. A., Lang, W. W., Maling, G. C., Nelson, D. E., Rader, C. M., & Welch, P. D. (1967). What is the fast fourier transform? *Proceedings of the IEEE*, *55*(10), 1664–1674.
- Cohen, L., Dehaene, S., Naccache, L., Lehéricy, S., Dehaene-Lambertz, G., Hénauff, M.-A., & Michel, F. (2000). The visual word form area: Spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain*, *123*(2), 291–307.
- Colby, C. L., Duhamel, J.-R., & Goldberg, M. E. (1996). Visual, presaccadic, and cognitive activation of single neurons in monkey lateral intraparietal area. *Journal of neurophysiology*, *76*(5), 2841–2852.
- Collin, C. A., Liu, C. H., Troje, N. F., McMullen, P. A., & Chaudhuri, A. (2004). Face recognition is affected by similarity in spatial frequency range to a greater degree than within-category object recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(5), 975.
- Collins, J. A., & Olson, I. R. (2014). Beyond the FFA: The role of the ventral anterior temporal lobes in face processing. *Neuropsychologia*, *61*, 65–79.
- Coogan, T., & Burkhalter, A. (1990). Conserved patterns of cortico-cortical connections define areal hierarchy in rat visual cortex. *Experimental brain research*, *80*, 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.
- Corthout, E., & Supèr, H. (2004). Contextual modulation in v1: The rossi-zipser controversy. *Experimental brain research*, *156*, 118–123.
- Corthout, E., Uttl, B., Walsh, V., Hallett, M., & Cowey, A. (1999). Timing of activity in early visual cortex as revealed by transcranial magnetic stimulation. *Neuroreport*, *10*(12), 2631–2634.
- Corthout, E., Uttl, B., Ziemann, U., Cowey, A., & Hallett, M. (1998). Two periods of processing in the (circum) striate visual cortex as revealed by transcranial magnetic stimulation. *Neuropsychologia*, *37*(2), 137–145.
- Costen, N. P., Parker, D. M., & Craw, I. (1994). Spatial content and spatial quantisation effects in face recognition. *Perception*, *23*(2), 129–146.
- Costen, N. P., Parker, D. M., & Craw, I. (1996). Effects of high-pass and low-pass spatial filtering on face identification. *Perception & psychophysics*, *58*, 602–612.
- Cox, R. W., & Hyde, J. S. (1997). Software tools for analysis and visualization of fmri data. *NMR in Biomedicine: An International Journal Devoted to the Development and Application of Magnetic Resonance In Vivo*, *10*(4–5), 171–178.
- Daelli, V., van Rijsbergen, N. J., & Treves, A. (2010). How recent experience affects the perception of ambiguous objects. *Brain research*, *1322*, 81–91.

- Dakin, S. C., & Watt, R. J. (2009). Biological bar codes in human faces. *Journal of vision*, 9(4), 2–2.
- Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis. i. segmentation and surface reconstruction. *Neuroimage*, 9(2), 179–194.
- Dalrymple, K. A., Oruc, I., Duchaine, B., Pancaroglu, R., Fox, C. J., Iaria, G., Handy, T. C., & Barton, J. J. (2011). The anatomic basis of the right face-selective n170 in acquired prosopagnosia: A combined erp/fmri study. *Neuropsychologia*, 49(9), 2553–2563.
- Davies, G., Ellis, H. D., & Shepherd, J. (1978). Face recognition accuracy as a function of mode of representation. *Journal of applied psychology*, 63(2), 180.
- Davies-Thompson, J., & Andrews, T. J. (2012). Intra-and interhemispheric connectivity between face-selective regions in the human brain. *Journal of neurophysiology*, 108(11), 3087–3095.
- De Cesarei, A., & Codispoti, M. (2013). Spatial frequencies and emotional perception. *Reviews in the Neurosciences*, 24(1), 89–104.
- 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.
- De Valois, R. L., & De Valois, K. K. (1980). Spatial vision. *Annual review of psychology*, 31(1), 309–341.
- DeAngelis, G. C., Ghose, G. M., Ohzawa, I., & Freeman, R. D. (1999). Functional micro-organization of primary visual cortex: Receptive field analysis of nearby neurons. *Journal of Neuroscience*, 19(10), 4046–4064.
- Deco, G., & Lee, T. S. (2004). The role of early visual cortex in visual integration: A neural model of recurrent interaction. *Eur. J. Neurosci.*, 20(4), 1089–1100.
- Dehaene, S., Naccache, L., Cohen, L., Bihan, D. L., Mangin, J.-F., Poline, J.-B., & Rivière, D. (2001). Cerebral mechanisms of word masking and unconscious repetition priming. *Nature neuroscience*, 4(7), 752–758.
- Delvenne, J.-F., Seron, X., Coyette, F., & Rossion, B. (2004). Evidence for perceptual deficits in associative visual (prosop) agnosia: A single-case study. *Neuropsychologia*, 42(5), 597–612.
- den Ouden, H. E., Daunizeau, J., Roiser, J., Friston, K. J., & Stephan, K. E. (2010). Striatal prediction error modulates cortical coupling. *Journal of Neuroscience*, 30(9), 3210–3219.
- DeValois, R. L., Burke, P., De Valois, K. K., & DeValois, K. K. (1990). *Spatial vision*. Oxford University Press on Demand.
- Devue, C., Collette, F., Balteau, E., Degueldre, C., Luxen, A., Maquet, P., & Brédart, S. (2007). Here i am: The cortical correlates of visual self-recognition. *Brain research*, 1143, 169–182.
- de-Wit, L. H., Kubilius, J., Wagemans, J., & Op de Beeck, H. P. (2012). Bistable gestalts reduce activity in the whole of v1, not just the retinotopically predicted parts. *J. Vis.*, 12(11).

- DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron*, 73(3), 415–434.
- Dobs, K., Bühlhoff, I., & Schultz, J. (2016). Identity information content depends on the type of facial movement. *Scientific reports*, 6(1), 34301.
- Dobs, K., Isik, L., Pantazis, D., & Kanwisher, N. (2019). How face perception unfolds over time. *Nature communications*, 10(1), 1258.
- Douglas, R. J., & Martin, K. A. (2004). Neuronal circuits of the neocortex. *Annu. Rev. Neurosci.*, 27, 419–451.
- Douglas, R. J., & Martin, K. A. (2007). Mapping the matrix: The ways of neocortex. *Neuron*, 56(2), 226–238.
- Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science*, 293(5539), 2470–2473.
- Dubner, R., & Zeki, S. (1971). Response properities and receptive fields of cells in an anatomically defined region of the superior temporal sulcus in the monkey. *Brain research*.
- Duffy, C. J., & Wurtz, R. H. (1991). Sensitivity of mst neurons to optic flow stimuli. ii. mechanisms of response selectivity revealed by small-field stimuli. *Journal of neurophysiology*, 65(6), 1346–1359.
- Edelman, G. M., & Gally, J. A. (2013). Reentry: A key mechanism for integration of brain function. *Front. Integr. Neurosci.*, 7, 63.
- Eick, C. M., Kovács, G., Rostalski, S.-M., Röhrig, L., & Ambrus, G. G. (2020). The occipital face area is causally involved in identity-related visual-semantic associations. *Brain Structure and Function*, 225, 1483–1493.
- Einevoll, G. T., Jurkus, P., & Heggelund, P. (2011). Coarse-to-fine changes of receptive fields in lateral geniculate nucleus have a transient and a sustained component that depend on distinct mechanisms. *PloS one*, 6(9), e24523.
- Enns, J. T., & Di Lollo. (2000). What’s new in visual masking? *Trends Cogn. Sci.*, 4(9), 345–352.
- Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, 392(6676), 598–601.
- Eriksson, D., & Roland, P. (2006). Feed-forward, feedback and lateral interactions in membrane potentials and spike trains from the visual cortex in vivo. *Journal of Physiology-Paris*, 100(1-3), 100–109.
- Esteban, O., Blair, R., Markiewicz, C. J., Berleant, S. L., Moodie, C., Ma, F., Isik, A. I., Erramuzpe, A., Kent, M., Goncalves, M., et al. (2018). Fmriprep. *Software*.
- Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., et al. (2019). Fmriprep: A robust preprocessing pipeline for functional mri. *Nature methods*, 16(1), 111–116.
- Esterman, M., & Yantis, S. (2010). Perceptual expectation evokes category-selective cortical activity. *Cerebral Cortex*, 20(5), 1245–1253.

- Everson, R., Prashanth, A., Gabbay, M., Knight, B., Sirovich, L., & Kaplan, E. (1998). Representation of spatial frequency and orientation in the visual cortex. *Proceedings of the National Academy of Sciences*, 95(14), 8334–8338.
- Fahrenfort, J. J., Scholte, H. S., & Lamme, V. A. (2007). Masking disrupts reentrant processing in human visual cortex. *Journal of cognitive neuroscience*, 19(9), 1488–1497.
- Fahrenfort, J. J., Snijders, T. M., Heinen, K., van Gaal, S., Scholte, H. S., & Lamme, V. A. F. (2012). Neuronal integration in visual cortex elevates face category tuning to conscious face perception. *Proc. Natl. Acad. Sci. U. S. A.*, 109(52), 21504–21509.
- Fairhall, S. L., & Ishai, A. (2007). Effective connectivity within the distributed cortical network for face perception. *Cerebral cortex*, 17(10), 2400–2406.
- Fan, X., Wang, F., Shao, H., Zhang, P., & He, S. (2020). The bottom-up and top-down processing of faces in the human occipitotemporal cortex. *Elife*, 9, e48764.
- Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral cortex (New York, NY: 1991)*, 1(1), 1–47.
- Felsen, G., & Dan, Y. (2005). A natural approach to studying vision. *Nature neuroscience*, 8(12), 1643–1646.
- Felsen, G., Touryan, J., Han, F., & Dan, Y. (2005). Cortical sensitivity to visual features in natural scenes. *PLoS biology*, 3(10), e342.
- 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.
- Finzi, D., Gomez, J., Nordt, M., Rezai, A. A., Poltoratski, S., & Grill-Spector, K. (2021). Differential spatial computations in ventral and lateral face-selective regions are scaffolded by structural connections. *Nat. Commun.*, 12(1), 2278.
- Fiorentini, A., Maffei, L., & Sandini, G. (1983). The role of high spatial frequencies in face perception. *Perception*, 12(2), 195–201.
- Foley, E., Rippon, G., Thai, N. J., Longe, O., & Senior, C. (2012). Dynamic facial expressions evoke distinct activation in the face perception network: A connectivity analysis study. *Journal of cognitive neuroscience*, 24(2), 507–520.
- Fox, C. J., Iaria, G., & Barton, J. J. S. (2009). Defining the face processing network: Optimization of the functional localizer in fMRI. *Hum. Brain Mapp.*, 30(5), 1637–1651.
- Foxe, J. J., & Simpson, G. V. (2002). Flow of activation from v1 to frontal cortex in humans: A framework for defining "early" visual processing. *Experimental brain research*, 142, 139–150.
- Francis, G. (1997). Cortical dynamics of lateral inhibition: Metacontrast masking. *Psychological review*, 104(3), 572.

- Frazor, R. A., Albrecht, D. G., Geisler, W. S., & Crane, A. M. (2004). Visual cortex neurons of monkeys and cats: Temporal dynamics of the spatial frequency response function. *Journal of Neurophysiology*, 91(6), 2607–2627.
- Freud, E., Culham, J. C., Plaut, D. C., & Behrmann, M. (2017). The large-scale organization of shape processing in the ventral and dorsal pathways. *elife*, 6, e27576.
- Friston, K. (2005). A theory of cortical responses. *Philosophical transactions of the Royal Society B: Biological sciences*, 360(1456), 815–836.
- Galletti, C., Fattori, P., Kutz, D. F., & Battaglini, P. P. (1997). Arm movement-related neurons in the visual area v6a of the macaque superior parietal lobule. *European Journal of Neuroscience*, 9(2), 410–413.
- Gao, Z., & Bentin, S. (2011). Coarse-to-fine encoding of spatial frequency information into visual short-term memory for faces but impartial decay. *J. Exp. Psychol. Hum. Percept. Perform.*, 37(4), 1051–1064.
- Gattass, R., Nascimento-Silva, S., Soares, J. G., Lima, B., Jansen, A. K., Diogo, A. C. M., Farias, M. F., Botelho, E. P., Marco Marcondes, Mariani, O. S., Azzi, J., et al. (2005). Cortical visual areas in monkeys: Location, topography, connections, columns, plasticity and cortical dynamics. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1456), 709–731.
- Gauthier, I., Tarr, M. J., Moylan, J., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). The fusiform face area is part of a network that processes faces at the individual level. *Journal of cognitive neuroscience*, 12(3), 495–504.
- Gazzaley, A., Cooney, J. W., McEvoy, K., Knight, R. T., & D’esposito, M. (2005). Top-down enhancement and suppression of the magnitude and speed of neural activity. *Journal of cognitive neuroscience*, 17(3), 507–517.
- Gentile, F., Ales, J., & Rossion, B. (2017). Being BOLD: The neural dynamics of face perception. *Hum. Brain Mapp.*, 38(1), 120–139.
- George, N., Dolan, R. J., Fink, G. R., Baylis, G. C., Russell, C., & Driver, J. (1999). Contrast polarity and face recognition in the human fusiform gyrus. *Nat. Neurosci.*, 2(6), 574–580.
- Gilaie-Dotan, S., & Malach, R. (2007). Sub-exemplar shape tuning in human face-related areas. *Cerebral cortex*, 17(2), 325–338.
- Gilbert, C. D., Hirsch, J. A., & Wiesel, T. N. (1990). Lateral interactions in visual cortex. *Cold Spring Harbor symposia on quantitative biology*, 55, 663–677.
- Gilbert, C. D., & Li, W. (2013). Top-down influences on visual processing. *Nature Reviews Neuroscience*, 14(5), 350–363.
- Gilbert, C. D., & Sigman, M. (2007). Brain states: Top-down influences in sensory processing. *Neuron*, 54(5), 677–696.
- Gilbert, C. D., & Wiesel, T. N. (1983). Clustered intrinsic connections in cat visual cortex. *Journal of Neuroscience*, 3(5), 1116–1133.
- Ginsburg, A. P. (1980). Specific relevant spatial information for image evaluation and display design: An explanation of how we see certain objects.

- Glover, G. H., Li, T.-Q., & Ress, D. (2000). Image-based method for retrospective correction of physiological motion effects in fmri: Retroicor. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 44(1), 162–167.
- Gobbini, M. I., & Haxby, J. V. (2006). Neural response to the visual familiarity of faces. *Brain research bulletin*, 71(1-3), 76–82.
- Gobbini, M. I., & Haxby, J. V. (2007). Neural systems for recognition of familiar faces. *Neuropsychologia*, 45(1), 32–41.
- Goebel, R., Muckli, L., & Kim, D.-S. (2004). Visual system. *The Human Nervous System Elsevier, San Diego*, 1280–1305.
- 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., Gauthier, I., & Rossion, B. (2003). Spatial scale contribution to early visual differences between face and object processing. *Cognitive Brain Research*, 16(3), 416–424.
- Goffaux, V., Hault, B., Michel, C., Vuong, Q. C., & Rossion, B. (2005). The respective role of low and high spatial frequencies in supporting configural and featural processing of faces. *Perception*, 34(1), 77–86.
- Goffaux, V., Peters, J., Haubrechts, J., Schiltz, C., Jansma, B., & Goebel, R. (2011). From coarse to fine? spatial and temporal dynamics of cortical face processing. *Cerebral Cortex*, 21(2), 467–476.
- Goffaux, V., Rossion, B., Sorger, B., Schiltz, C., & Goebel, R. (2009). Face inversion disrupts the perception of vertical relations between features in the right human occipito-temporal cortex. *Journal of Neuropsychology*, 3(1), 45–67.
- Goffaux, V., Schiltz, C., Mur, M., & Goebel, R. (2012). Local discriminability determines the strength of holistic processing for faces in the fusiform face area. *Front. Psychol.*, 3, 604.
- Goffaux, V., Schiltz, C., Mur, M., & Goebel, R. (2013). Local discriminability determines the strength of holistic processing for faces in the fusiform face area. *Frontiers in psychology*, 3, 604.
- Goffaux, V., & Rossion, B. (2006). Faces are "spatial"—holistic face perception is supported by low spatial frequencies. *Journal of Experimental Psychology: Human perception and performance*, 32(4), 1023.
- Gold, J., Bennett, P. J., & Sekuler, A. B. (1999). Identification of band-pass filtered letters and faces by human and ideal observers. *Vision research*, 39(21), 3537–3560.
- Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in neurosciences*, 15(1), 20–25.
- Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., Flandin, G., Ghosh, S. S., Glatard, T., Halchenko, Y. O., Handwerker, D. A., Hanke, M., Keator, D., Li, X., Michael, Z., Maumet, C., Nichols, B. N., Nichols, T. E., Pellman, J., ... Poldrack, R. A. (2016).

- The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Sci. Data*, 3, 160044.
- Gorgolewski, K. J., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., & Ghosh, S. S. (2011). Nipype: A flexible, lightweight and extensible neuroimaging data processing framework in python. *Frontiers in neuroinformatics*, 13.
- Gorgolewski, K. J., Nichols, T., Kennedy, D. N., Poline, J.-B., & Poldrack, R. A. (2018). Making replication prestigious. *Behav. Brain Sci.*, 41(e131), e131.
- Graham, N. V. S. (1989). *Visual pattern analyzers*. Oxford University Press.
- Gras, V., Vignaud, A., Amadon, A., Le Bihan, D., & Boulant, N. (2017). Universal pulses: A new concept for calibration-free parallel transmission. *Magnetic resonance in medicine*, 77(2), 635–643.
- Green, M. F., Glahn, D., Engel, S. A., Nuechterlein, K. H., Sabb, F., Strojwas, M., & Cohen, M. S. (2005). Regional brain activity associated with visual backward masking. *J. Cogn. Neurosci.*, 17(1), 13–23.
- Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., & Altman, D. G. (2016). Statistical tests, p values, confidence intervals, and power: A guide to misinterpretations. *European journal of epidemiology*, 31, 337–350.
- Greve, D. N., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *Neuroimage*, 48(1), 63–72.
- Grill-Spector, K., Kushnir, T., Hendler, T., & Malach, R. (2000). The dynamics of object-selective activation correlate with recognition performance in humans. *Nat. Neurosci.*, 3(8), 837–843.
- Grill-Spector, K., & Kanwisher, N. (2005). Visual recognition: As soon as you know it is there, you know what it is. *Psychological Science*, 16(2), 152–160.
- Grill-Spector, K., & Weiner, K. S. (2014). The functional architecture of the ventral temporal cortex and its role in categorization. *Nature Reviews Neuroscience*, 15(8), 536–548.
- Grill-Spector, K., Weiner, K. S., Gomez, J., Stigliani, A., & Natu, V. S. (2018). The functional neuroanatomy of face perception: From brain measurements to deep neural networks. *Interface Focus*, 8(4), 20180013.
- Grill-Spector, K., Weiner, K. S., Kay, K., & Gomez, J. (2017). The functional neuroanatomy of human face perception. *Annu. Rev. Vis. Sci.*, 3, 167–196.
- Grimaldi, P., Saleem, K. S., & Tsao, D. (2016). Anatomical connections of the functionally defined face patches in the macaque monkey. *Neuron*, 90(6), 1325–1342.
- Groen, I. I., Jahfari, S., Seijdel, N., Ghebreab, S., Lamme, V. A., & Scholte, H. S. (2018). Scene complexity modulates degree of feedback activity during object detection in natural scenes. *PLoS computational biology*, 14(12), e1006690.

- Gschwind, M., Pourtois, G., Schwartz, S., Van De Ville, D., & Vuilleumier, P. (2012). White-matter connectivity between face-responsive regions in the human brain. *Cerebral cortex*, 22(7), 1564–1576.
- Halekoh, U., & Højsgaard, S. (2014). A kenward-roger approximation and parametric bootstrap methods for tests in linear mixed models—the r package pbkrtest. *Journal of Statistical Software*, 59, 1–32.
- Halgren, E., Dale, A. M., Sereno, M. I., Tootell, R. B., Marinkovic, K., & Rosen, B. R. (1999). Location of human face-selective cortex with respect to retinotopic areas. *Human brain mapping*, 7(1), 29–37.
- 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 Res.*, 1117(1), 154–161.
- Hancock, P. J., Bruce, V., & Burton, A. M. (2000). Recognition of unfamiliar faces. *Trends in cognitive sciences*, 4(9), 330–337.
- Hansen, B. C., Greene, M. R., & Field, D. J. (2021). Dynamic electrode-to-image (deti) mapping reveals the human brains spatiotemporal code of visual information. *PLoS Computational Biology*, 17(9), e1009456.
- Hansen, B. C., Thompson, B., Hess, R. F., & Ellemberg, D. (2010). Extracting the internal representation of faces from human brain activity: An analogue to reverse correlation. *NeuroImage*, 51(1), 373–390.
- Harmon, L. D. (1973). The recognition of faces. *Scientific American*, 229(5), 70–83.
- Harmon, L. D., & Julesz, B. (1973). Masking in visual recognition: Effects of two-dimensional filtered noise. *Science*, 180(4091), 1194–1197.
- Harrison, L. M., Stephan, K. E., Rees, G., & Friston, K. J. (2007). Extra-classical receptive field effects measured in striate cortex with fmri. *Neuroimage*, 34(3), 1199–1208.
- Hartline, H. K. (1938). The response of single optic nerve fibers of the vertebrate eye to illumination of the retina. *American Journal of Physiology-Legacy Content*, 121(2), 400–415.
- Harvey, A. K., Pattinson, K. T., Brooks, J. C., Mayhew, S. D., Jenkinson, M., & Wise, R. G. (2008). Brainstem functional magnetic resonance imaging: Disentangling signal from physiological noise. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 28(6), 1337–1344.
- Harvey, M. A., Valentiniene, S., & Roland, P. E. (2009). Cortical membrane potential dynamics and laminar firing during object motion. *Front. Syst. Neurosci.*, 3, 7.
- Hawken, M. J., Shapley, R. M., Disney, A. A., Garcia-Marin, V., Henrie, A., Henry, C. A., Johnson, E. N., Joshi, S., Kelly, J. G., Ringach, D. L., et al. (2020). Functional clusters of neurons in layer 6 of macaque v1. *Journal of Neuroscience*, 40(12), 2445–2457.
- Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in cognitive sciences*, 4(6), 223–233.

- Hayes, A. (1988). Identification of two-tone images; some implications for high- and low-spatial-frequency processes in human vision. *Perception*, *17*(4), 429–436.
- Haynes, J.-D., Driver, J., & Rees, G. (2005). Visibility reflects dynamic changes of effective connectivity between v1 and fusiform cortex. *Neuron*, *46*(5), 811–821.
- Heekeren, H. R., Marrett, S., Bandettini, P. A., & Ungerleider, L. G. (2004). A general mechanism for perceptual decision-making in the human brain. *Nature*, *431*(7010), 859–862.
- Hegd , J. (2008). Time course of visual perception: Coarse-to-fine processing and beyond. *Progress in neurobiology*, *84*(4), 405–439.
- Heinze, G., Wallisch, C., & Dunkler, D. (2018). Variable selection - a review and recommendations for the practicing statistician. *Biom. J.*, *60*(3), 431–449.
- Hendry, S. H., & Yoshioka, T. (1994). A neurochemically distinct third channel in the macaque dorsal lateral geniculate nucleus. *Science*, *264*(5158), 575–577.
- Hermann, P., Bank ,  . M., G l, V., & Vidny nszky, Z. (2015). Neural basis of identity information extraction from noisy face images. *Journal of Neuroscience*, *35*(18), 7165–7173.
- Hill, H., & Johnston, A. (2001). Categorizing sex and identity from the biological motion of faces. *Current biology*, *11*(11), 880–885.
- Horovitz, S. G., Rossion, B., Skudlarski, P., & Gore, J. C. (2004). Parametric design and correlational analyses help integrating fmri and electrophysiological data during face processing. *Neuroimage*, *22*(4), 1587–1595.
- Hr ng-Tai Fai, A., & Cornelius, P. L. (1996). Approximate f-tests of multiple degree of freedom hypotheses in generalized least squares analyses of unbalanced split-plot experiments. *Journal of statistical computation and simulation*, *54*(4), 363–378.
- Hsiao, J. H.-w., & Cottrell, G. (2008). Two fixations suffice in face recognition. *Psychological science*, *19*(10), 998–1006.
- Hubel, D. H. (1960). Single unit activity in lateral geniculate body and optic tract of unrestrained cats. *The Journal of Physiology*, *150*(1), 91.
- Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex. *The Journal of physiology*, *160*(1), 106.
- Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of physiology*, *195*(1), 215–243.
- Hubel, D. H., & Wiesel, T. N. (1972). Laminar and columnar distribution of geniculocortical fibers in the macaque monkey. *Journal of Comparative Neurology*, *146*(4), 421–450.
- Hubel, D. H., Wiesel, T. N., & Stryker, M. P. (1978). Anatomical demonstration of orientation columns in macaque monkey. *Journal of Comparative Neurology*, *177*(3), 361–379.

- Hübener, M., Shoham, D., Grinvald, A., & Bonhoeffer, T. (1997). Spatial relationships among three columnar systems in cat area 17. *Journal of Neuroscience*, 17(23), 9270–9284.
- Hughes, H. C., Nozawa, G., & Kitterle, F. (1996). Global precedence, spatial frequency channels, and the statistics of natural images. *J. Cogn. Neurosci.*, 8(3), 197–230.
- Hupé, J., James, A., Payne, B., Lomber, S., Girard, P., & Bullier, J. (1998). Cortical feedback improves discrimination between figure and background by v1, v2 and v3 neurons. *Nature*, 394(6695), 784–787.
- Hutton, C., Josephs, O., Stadler, J., Featherstone, E., Reid, A., Speck, O., Bernarding, J., & Weiskopf, N. (2011). The impact of physiological noise correction on fmri at 7 t. *Neuroimage*, 57(1), 101–112.
- Ishai, A. (2008). Lets face it: Itsa cortical network. *Neuroimage*, 40(2), 415–419.
- 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.
- Issa, E. B., Cadieu, C. F., & DiCarlo, J. J. (2018). Neural dynamics at successive stages of the ventral visual stream are consistent with hierarchical error signals. *Elife*, 7, e42870.
- Issa, N. P., Trepel, C., & Stryker, M. P. (2000). Spatial frequency maps in cat visual cortex. *Journal of Neuroscience*, 20(22), 8504–8514.
- Ito, M., & Gilbert, C. D. (1999). Attention modulates contextual influences in the primary visual cortex of alert monkeys. *Neuron*, 22(3), 593–604.
- 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.
- Jehee, J. F., Roelfsema, P. R., Deco, G., Murre, J. M., & Lamme, V. A. (2007). Interactions between higher and lower visual areas improve shape selectivity of higher level neuronsexplaining crowding phenomena. *Brain Research*, 1157, 167–176.
- Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*, 17(2), 825–841.
- Jiang, X., Rosen, E., Zeffiro, T., VanMeter, J., Blanz, V., & Riesenhuber, M. (2006). Evaluation of a shape-based model of human face discrimination using fmri and behavioral techniques. *Neuron*, 50(1), 159–172.
- Jones, R., & Keck, M. J. (1978). Visual evoked response as a function of grating spatial frequency. *Invest. Ophthalmol. Vis. Sci.*, 17(7), 652–659.
- Kadosh, K. C., Walsh, V., & Kadosh, R. C. (2011). Investigating face-property specific processing in the right ofa. *Social cognitive and affective neuroscience*, 6(1), 58–65.
- Kanizsa, G. (1976). Subjective contours. *Scientific American*, 234(4), 48–53.
- Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of neuroscience*, 17(11), 4302–4311.

- Kaplan, E., & Shapley, R. M. (1986). The primate retina contains two types of ganglion cells, with high and low contrast sensitivity. *Proceedings of the National Academy of Sciences*, 83(8), 2755–2757.
- Kar, K., & DiCarlo, J. J. (2021). Fast recurrent processing via ventrolateral prefrontal cortex is needed by the primate ventral stream for robust core visual object recognition. *Neuron*, 109(1), 164–176.e5.
- 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. *Nat. Neurosci.*, 22(6), 974–983.
- Kasper, L., Bollmann, S., Diaconescu, A. O., Hutton, C., Heinzle, J., Iglesias, S., Hauser, T. U., Sebold, M., Manjaly, Z.-M., Pruessmann, K. P., et al. (2017). The physio toolbox for modeling physiological noise in fmri data. *Journal of neuroscience methods*, 276, 56–72.
- Kauffmann, L., Bourgin, J., Guyader, N., & Peyrin, C. (2015). The neural bases of the semantic interference of spatial frequency-based information in scenes. *J. Cogn. Neurosci.*, 27(12), 2394–2405.
- Kauffmann, L., Chauvin, A., Guyader, N., & Peyrin, C. (2015). Rapid scene categorization: Role of spatial frequency order, accumulation mode and luminance contrast. *Vision Res.*, 107, 49–57.
- Kauffmann, L., Chauvin, A., Pichat, C., & Peyrin, C. (2015). Effective connectivity in the neural network underlying coarse-to-fine categorization of visual scenes. a dynamic causal modeling study. *Brain Cogn.*, 99, 46–56.
- Kauffmann, L., Ramanoël, S., Guyader, N., Chauvin, A., & Peyrin, C. (2015). Spatial frequency processing in scene-selective cortical regions. *NeuroImage*, 112, 86–95.
- Kauffmann, L., Roux-Sibilon, A., Beffara, B., Mermillod, M., Guyader, N., & Peyrin, C. (2017). How does information from low and high spatial frequencies interact during scene categorization? *Visual Cognition*, 25(9-10), 853–867.
- Kay, K. N., & Yeatman, J. D. (2016). Bottom-up and top-down computations in high-level visual cortex. *BioRxiv*, 053595.
- Kazandjian, T. D., Petras, D., Robinson, S. D., van Thiel, J., Greene, H. W., Arbuckle, K., Barlow, A., Carter, D. A., Wouters, R. M., Whiteley, G., et al. (2021). Convergent evolution of pain-inducing defensive venom components in spitting cobras. *Science*, 371(6527), 386–390.
- Keesey, U. T. (1972). Flicker and pattern detection: A comparison of thresholds. *JOSA*, 62(3), 446–448.
- Keller, G. B., & Mrsic-Flogel, T. D. (2018). Predictive processing: A canonical cortical computation. *Neuron*, 100(2), 424–435.
- Kenward, M. G., & Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, 983–997.
- Keysers, C., Xiao, D. K., Földiák, P., & Perrett, D. I. (2001). The speed of sight. *J. Cogn. Neurosci.*, 13(1), 90–101.
- Keysers, C., & Perrett, D. I. (2002). Visual masking and RSVP reveal neural competition. *Trends Cogn. Sci.*, 6(3), 120–125.

- Klein, A., Ghosh, S. S., Bao, F. S., Giard, J., Häme, Y., Stavsky, E., Lee, N., Rossa, B., Reuter, M., Chaibub Neto, E., & Keshavan, A. (2017). Mind-boggling morphometry of human brains. *PLoS Comput. Biol.*, *13*(2), e1005350.
- 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.
- Koivisto, M., Railo, H., Revonsuo, A., Vanni, S., & Salminen-Vaparanta, N. (2011). Recurrent processing in v1/v2 contributes to categorization of natural scenes. *Journal of Neuroscience*, *31*(7), 2488–2492.
- Kok, P., & de Lange, F. P. (2014). Shape perception simultaneously up- and downregulates neural activity in the primary visual cortex. *Curr. Biol.*, *24*(13), 1531–1535.
- 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., & De Lange, F. P. (2012). Less is more: Expectation sharpens representations in the primary visual cortex. *Neuron*, *75*(2), 265–270.
- Kolb, H., Fernandez, E., & Nelson, R. (1995). Webvision: The organization of the retina and visual system [internet].
- Komatsu, H., & Wurtz, R. H. (1988). Relation of cortical areas mt and mst to pursuit eye movements. iii. interaction with full-field visual stimulation. *Journal of Neurophysiology*, *60*(2), 621–644.
- Kornowski, J. A., & Petersik, J. T. (2003). Effects on face recognition of spatial-frequency information contained in inspection and test stimuli. *The Journal of general psychology*, *130*(3), 229–244.
- Kovács, G., Vogels, R., & Orban, G. A. (1995). Cortical correlate of pattern backward masking. *Proc. Natl. Acad. Sci. U. S. A.*, *92*(12), 5587–5591.
- Kramer, R. S., Young, A. W., & Burton, A. M. (2018). Understanding face familiarity. *Cognition*, *172*, 46–58.
- Krause, M. R., & Pack, C. C. (2014). Contextual modulation and stimulus selectivity in extrastriate cortex. *Vision Research*, *104*, 36–46.
- Kravitz, D. J., Saleem, K. S., Baker, C. I., Ungerleider, L. G., & Mishkin, M. (2013). The ventral visual pathway: An expanded neural framework for the processing of object quality. *Trends in cognitive sciences*, *17*(1), 26–49.
- Kreiman, G., & Serre, T. (2020). Beyond the feedforward sweep: Feedback computations in the visual cortex. *Annals of the New York Academy of Sciences*, *1464*(1), 222–241.
- Kulikowski, J., & Tolhurst, D. (1973). Psychophysical evidence for sustained and transient detectors in human vision. *The Journal of Physiology*, *232*(1), 149–162.
- 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.

- Kwong, K. K., Belliveau, J. W., Chesler, D. A., Goldberg, I. E., Weisskoff, R. M., Poncelet, B. P., Kennedy, D. N., Hoppel, B. E., Cohen, M. S., & Turner, R. (1992). Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proceedings of the National Academy of Sciences*, 89(12), 5675–5679.
- Laguesse, R., Dormal, G., Biervoye, A., Kuefner, D., & Rossion, B. (2012). Extensive visual training in adulthood significantly reduces the face inversion effect. *Journal of vision*, 12(10), 14–14.
- Lamme, V. A. F. (1995). The neurophysiology of figure-ground segregation in primary visual cortex. *Journal of neuroscience*, 15(2), 1605–1615.
- Lamme, V. A. F., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci.*, 23(11), 571–579.
- Lamme, V. A. F., Super, H., & Spekreijse, H. (1998). Feedforward, horizontal, and feedback processing in the visual cortex. *Current opinion in neurobiology*, 8(4), 529–535.
- Lamme, V. A. F., Zipser, K., & Spekreijse, H. (2002). Masking interrupts figure-ground signals in V1. *J. Cogn. Neurosci.*, 14(7), 1044–1053.
- Landi, S. M., & Freiwald, W. A. (2017). Two areas for familiar face recognition in the primate brain. *Science*, 357(6351), 591–595.
- Lee, T. S. (2015). The visual system’s internal model of the world. *Proc. IEEE Inst. Electr. Electron. Eng.*, 103(8), 1359–1378.
- Lee, T. S., Mumford, D., Romero, R., & Lamme, V. A. (1998). The role of the primary visual cortex in higher level vision. *Vision Res.*, 38(15-16), 2429–2454.
- Lee, T. S., & Mumford, D. (2003). Hierarchical bayesian inference in the visual cortex. *JOSA A*, 20(7), 1434–1448.
- Lee, T. S., & Nguyen, M. (2001). Dynamics of subjective contour formation in the early visual cortex. *Proceedings of the National Academy of Sciences*, 98(4), 1907–1911.
- Lee, T. S., Yang, C. F., Romero, R. D., & Mumford, D. (2002). Neural activity in early visual cortex reflects behavioral experience and higher-order perceptual saliency. *Nature neuroscience*, 5(6), 589–597.
- Lee, Y., Matsumiya, K., & Wilson, H. R. (2006). Size-invariant but viewpoint-dependent representation of faces. *Vision research*, 46(12), 1901–1910.
- Leek, M. R. (2001). Adaptive procedures in psychophysical research. *Perception & psychophysics*, 63(8), 1279–1292.
- Lerner, Y., Hendler, T., Ben-Bashat, D., Harel, M., & Malach, R. (2001). A hierarchical axis of object processing stages in the human visual cortex. *Cerebral cortex*, 11(4), 287–297.
- Leveroni, C. L., Seidenberg, M., Mayer, A. R., Mead, L. A., Binder, J. R., & Rao, S. M. (2000). Neural systems underlying the recognition of familiar and newly learned faces. *Journal of Neuroscience*, 20(2), 878–886.

- Levy, I., Hasson, U., Avidan, G., Hendler, T., & Malach, R. (2001). Center-periphery organization of human object areas. *Nature neuroscience*, 4(5), 533–539.
- Liu, C. H., Collin, C. A., Rainville, S. J., & Chaudhuri, A. (2000). The effects of spatial frequency overlap on face recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 26(3), 956.
- Liu, J., Harris, A., & Kanwisher, N. (2013). Stages of processing in face perception: An meg study. In *Social neuroscience* (pp. 75–85). Psychology Press.
- Liu, J., Li, J., Feng, L., Li, L., Tian, J., & Lee, K. (2014). Seeing Jesus in toast: Neural and behavioral correlates of face pareidolia. *Cortex*, 53, 60–77.
- Liu-Shuang, J., Yang, Y.-F., Rossion, B., & Goffaux, V. (2022). Natural contrast statistics facilitate human face categorization. *Eneuro*.
- Livingstone, M., & Hubel, D. (1988). Segregation of form, color, movement, and depth: Anatomy, physiology, and perception. *Science*, 240(4853), 740–749.
- Lochmann, T., & Deneve, S. (2011). Neural processing as causal inference. *Current opinion in neurobiology*, 21(5), 774–781.
- Loftus, G. R., & Masson, M. E. (1994). Using confidence intervals in within-subject designs. *Psychon. Bull. Rev.*, 1(4), 476–490.
- Lu, Y., Yin, J., Chen, Z., Gong, H., Liu, Y., Qian, L., Li, X., Liu, R., Andolina, I. M., & Wang, W. (2018). Revealing detail along the visual hierarchy: Neural clustering preserves acuity from v1 to v4. *Neuron*, 98(2), 417–428.
- Luck, S. J., Chelazzi, L., Hillyard, S. A., & Desimone, R. (1997). Neural mechanisms of spatial selective attention in areas v1, v2, and v4 of macaque visual cortex. *Journal of neurophysiology*, 77(1), 24–42.
- MaBouDi, H., Shimazaki, H., Amari, S.-i., & Soltanian-Zadeh, H. (2016). Representation of higher-order statistical structures in natural scenes via spatial phase distributions. *Vision research*, 120, 61–73.
- Macknik, S. L., & Livingstone, M. S. (1998). Neuronal correlates of visibility and invisibility in the primate visual system. *Nature neuroscience*, 1(2), 144–149.
- Maffei, L., & Fiorentini, A. (1977). Spatial frequency rows in the striate visual cortex. *Vision research*, 17(2), 257–264.
- Malach, R., Reppas, J., Benson, R., Kwong, K., Jiang, H., Kennedy, W., Ledden, P., Brady, T., Rosen, B., & Tootell, R. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proceedings of the National Academy of Sciences*, 92(18), 8135–8139.
- Malone, B. J., Kumar, V. R., & Ringach, D. L. (2007). Dynamics of receptive field size in primary visual cortex. *Journal of neurophysiology*, 97(1), 407–414.
- Maniglia, M., Trotter, Y., & Aedo-Jury, F. (2019). Tms reveals inhibitory extrastriate cortico-cortical feedback modulation of v1 activity in humans. *Brain Structure and Function*, 224(9), 3399–3408.

- Markov, N. T., Misery, P., Falchier, A., Lamy, C., Vezoli, J., Quilodran, R., Gariel, M., Giroud, P., Ercsey-Ravasz, M., Pilaz, L., et al. (2011). Weight consistency specifies regularities of macaque cortical networks. *Cerebral cortex*, 21(6), 1254–1272.
- Markov, N. T., Vezoli, J., Chameau, P., Falchier, A., Quilodran, R., Huissoud, C., Lamy, C., Misery, P., Giroud, P., Ullman, S., Barone, P., Dehay, C., Knoblauch, K., & Kennedy, H. (2014). Anatomy of hierarchy: Feedforward and feedback pathways in macaque visual cortex. *J. Comp. Neurol.*, 522(1), 225–259.
- Marques, J. P., Kober, T., Krueger, G., van der Zwaag, W., Van de Moortele, P.-F., & Gruetter, R. (2010). Mp2rage, a self bias-field corrected sequence for improved segmentation and t1-mapping at high field. *Neuroimage*, 49(2), 1271–1281.
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. Henry Holt; Co., Inc.
- Martnez, A., DiRusso, F., Anllo-Vento, L., Sereno, M. I., Buxton, R. B., & Hillyard, S. A. (2001). Putting spatial attention on the map: Timing and localization of stimulus selection processes in striate and extrastriate visual areas. *Vision research*, 41(10-11), 1437–1457.
- Mazer, J. A., Vinje, W. E., McDermott, J., Schiller, P. H., & Gallant, J. L. (2002). Spatial frequency and orientation tuning dynamics in area V1. *Proc. Natl. Acad. Sci. U. S. A.*, 99(3), 1645–1650.
- McCarthy, G., Puce, A., Gore, J. C., & Allison, T. (1997). Face-specific processing in the human fusiform gyrus. *Journal of cognitive neuroscience*, 9(5), 605–610.
- McGuire, B. A., Gilbert, C. D., Rivlin, P. K., & Wiesel, T. N. (1991). Targets of horizontal connections in macaque primary visual cortex. *Journal of Comparative Neurology*, 305(3), 370–392.
- Megreya, A. M., & Burton, A. M. (2006). Unfamiliar faces are not faces: Evidence from a matching task. *Memory & cognition*, 34, 865–876.
- Mehta, A. D., Ulbert, I., & Schroeder, C. E. (2000a). Intermodal selective attention in monkeys. i: Distribution and timing of effects across visual areas. *Cerebral cortex*, 10(4), 343–358.
- Mehta, A. D., Ulbert, I., & Schroeder, C. E. (2000b). Intermodal selective attention in monkeys. ii: Physiological mechanisms of modulation. *Cerebral Cortex*, 10(4), 359–370.
- Meister, M., & Berry, M. J. (1999). The neural code of the retina. *neuron*, 22(3), 435–450.
- Merigan, W. H., & Maunsell, J. H. (1993). How parallel are the primate visual pathways? *Annual review of neuroscience*, 16(1), 369–402.
- Merkel, D., et al. (2014). Docker: Lightweight linux containers for consistent development and deployment. *Linux j*, 239(2), 2.
- Mermillod, M., Bourrier, Y., David, E., Kauffmann, L., Chauvin, A., Guyader, N., Dutheil, F., & Peyrin, C. (2019). The importance of recurrent top-down

- synaptic connections for the anticipation of dynamic emotions. *Neural Networks*, 109, 19–30.
- 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.
- Mihaylova, M., Stomonyakov, V., & Vassilev, A. (1999). Peripheral and central delay in processing high spatial frequencies: Reaction time and VEP latency studies. *Vision Res.*, 39(4), 699–705.
- Miller, B. T., & D’Esposito, M. (2005). Searching for the top in top-down control. *Neuron*, 48(4), 535–538.
- Milner, A. D., & Goodale, M. (1995). *The visual brain in action*. Oxford University Press. <https://books.google.be/books?id=kPVqAAAAMAAJ>
- Milner, A. D., & Goodale, M. A. (2008). Two visual systems re-viewed. *Neuropsychologia*, 46(3), 774–785.
- Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: Two cortical pathways. *Trends in neurosciences*, 6, 414–417.
- Moeller, S., Freiwald, W. A., & Tsao, D. Y. (2008). Patches with links: A unified system for processing faces in the macaque temporal lobe. *Science*, 320(5881), 1355–1359.
- Mohsenzadeh, Y., Qin, S., Cichy, R. M., & Pantazis, D. (2018). Ultra-rapid serial visual presentation reveals dynamics of feedforward and feedback processes in the ventral visual pathway. *Elife*, 7, e36329.
- Monti, M. M. (2011). Statistical analysis of fmri time-series: A critical review of the glm approach. *Frontiers in human neuroscience*, 5, 28.
- Motter, B. C. (1993). Focal attention produces spatially selective processing in visual cortical areas v1, v2, and v4 in the presence of competing stimuli. *Journal of neurophysiology*, 70(3), 909–919.
- Mu, T., & Li, S. (2013). The neural signature of spatial frequency-based information integration in scene perception. *Experimental brain research*, 227, 367–377.
- Muckli, L. (2010). What are we missing here? brain imaging evidence for higher cognitive functions in primary visual cortex V1. *Int. J. Imaging Syst. Technol.*, 20(2), 131–139.
- Muckli, L., De Martino, F., Vizioli, L., Petro, L. S., Smith, F. W., Ugurbil, K., Goebel, R., & Yacoub, E. (2015). Contextual feedback to superficial layers of v1. *Current Biology*, 25(20), 2690–2695.
- Muckli, L., & Petro, L. S. (2013). Network interactions: Non-geniculate input to V1. *Curr. Opin. Neurobiol.*, 23(2), 195–201.
- Mumford, D. (1992). On the computational architecture of the neocortex. *Biological cybernetics*, 66(3), 241–251.
- Murray, S. O., Kersten, D., Olshausen, B. A., Schrater, P., & Woods, D. L. (2002). Shape perception reduces activity in human primary visual cortex. *Proceedings of the National Academy of Sciences*, 99(23), 15164–15169.

- Murray, S. O., & Wojciulik, E. (2004). Attention increases neural selectivity in the human lateral occipital complex. *Nature neuroscience*, 7(1), 70–74.
- 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.
- Musselwhite, M. J., & Jeffreys, D. A. (1985). The influence of spatial frequency on the reaction times and evoked potentials recorded to grating pattern stimuli. *Vision Res.*, 25(11), 1545–1555.
- Näsänen, R. (1999). Spatial frequency bandwidth used in the recognition of facial images. *Vision research*, 39(23), 3824–3833.
- Newsome, W. T., & Pare, E. B. (1988). A selective impairment of motion perception following lesions of the middle temporal visual area (mt). *Journal of Neuroscience*, 8(6), 2201–2211.
- Nielsen, K. J., Logothetis, N. K., & Rainer, G. (2006). Dissociation between local field potentials and spiking activity in macaque inferior temporal cortex reveals diagnosticity-based encoding of complex objects. *Journal of Neuroscience*, 26(38), 9639–9645.
- Nirody, J. A. (2014). Development of spatial coarse-to-fine processing in the visual pathway. *J. Comput. Neurosci.*, 36(3), 401–414.
- Noesselt, T., Hillyard, S. A., Woldorff, M. G., Schoenfeld, A., Hagner, T., Jäncke, L., Tempelmann, C., Hinrichs, H., & Heinze, H.-J. (2002). Delayed striate cortical activation during spatial attention. *Neuron*, 35(3), 575–587.
- Noudoost, B., & Moore, T. (2011). Control of visual cortical signals by prefrontal dopamine. *Nature*, 474(7351), 372–375.
- Nowak, L., Munk, M., Girard, P., & Bullier, J. (1995). Visual latencies in areas v1 and v2 of the macaque monkey. *Visual neuroscience*, 12(2), 371–384.
- Ogawa, S., Menon, R., Tank, D. W., Kim, S., Merkle, H., Ellermann, J., & Ugurbil, K. (1993). Functional brain mapping by blood oxygenation level-dependent contrast magnetic resonance imaging. a comparison of signal characteristics with a biophysical model. *Biophysical journal*, 64(3), 803–812.
- Ogawa, S., Tank, D. W., Menon, R., Ellermann, J. M., Kim, S. G., Merkle, H., & Ugurbil, K. (1992). Intrinsic signal changes accompanying sensory stimulation: Functional brain mapping with magnetic resonance imaging. *Proceedings of the National Academy of Sciences*, 89(13), 5951–5955.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The edinburgh inventory. *Neuropsychologia*, 9(1), 97–113.
- Olshausen, B. A., & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381(6583), 607–609.
- Olshausen, B. A., & Field, D. J. (2006, January). What is the other 85 percent of V1 doing? In *23 problems in systems neuroscience* (pp. 182–212). Oxford University Press New York.

- Op De Beeck, H., & Vogels, R. (2000). Spatial sensitivity of macaque inferior temporal neurons. *Journal of Comparative Neurology*, 426(4), 505–518.
- O'Reilly, J. X., Woolrich, M. W., Behrens, T. E. J., Smith, S. M., & Johansen-Berg, H. (2012). Tools of the trade: Psychophysiological interactions and functional connectivity. *Soc. Cogn. Affect. Neurosci.*, 7(5), 604–609.
- O'Reilly, R. C., Wyatte, D., Herd, S., Mingus, B., & Jilk, D. J. (2013). Recurrent processing during object recognition. *Frontiers in psychology*, 4, 124.
- Oruç, ., & Barton, J. J. (2010). Critical frequencies in the perception of letters, faces, and novel shapes: Evidence for limited scale invariance for faces. *Journal of Vision*, 10(12), 20–20.
- Osterberg, G. A. (1935). Topography of the layer of rods and cones in the human retina. *Acta ophthalmologica*.
- O'Toole, A. J., Roark, D. A., & Abdi, H. (2002). Recognizing moving faces: A psychological and neural synthesis. *Trends in cognitive sciences*, 6(6), 261–266.
- Pantle, A. (1970). Adaptation to pattern spatial frequency: Effects on visual movement sensitivity in humans. *JOSA*, 60(8), 1120–1124.
- Parker, D. M. (1980). Simple reaction times to the onset, offset, and contrast reversal of sinusoidal grating stimuli. *Percept. Psychophys.*, 28(4), 365–368.
- Parker, D. M., & Costen, N. P. (1999). One extreme or the other or perhaps the golden mean? issues of spatial resolution in face processing. *Current Psychology*, 18, 118–127.
- Parker, D. M., & Dutch, S. (1987). Perceptual latency and spatial frequency. *Vision Res.*, 27(8), 1279–1283.
- Parker, D. M., Lishman, J. R., & Hughes, J. (1997). Evidence for the view that temporospatial integration in vision is temporally anisotropic. *Perception*, 26(9), 1169–1180.
- Parker, D. M., Lishman, J. R., & Hughes, J. (1992). Temporal integration of spatially filtered visual images. *Perception*, 21(2), 147–160.
- Parker, D. M., & Salzen, E. (1977). Latency changes in the human visual evoked response to sinusoidal gratings. *Vision research*, 17(10), 1201–1204.
- Parker, D. M., & Salzen, E. A. (1977). The spatial selectivity of early and late waves within the human visual evoked response. *Perception*, 6(1), 85–95.
- Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 51(1), 195–203. <https://doi.org/10.3758/s13428-018-01193-y>
- Pennartz, C. M., Dora, S., Muckli, L., & Lorteije, J. A. (2019). Towards a unified view on pathways and functions of neural recurrent processing. *Trends in neurosciences*, 42(9), 589–603.
- Peters, J. C., Goebel, R., & Goffaux, V. (2018). From coarse to fine: Interactive feature processing precedes local feature analysis in human face perception. *Biological psychology*, 138, 1–10.

- Petras, K., Ten Oever, S., Dalal, S. S., & Goffaux, V. (2021). Information redundancy across spatial scales modulates early visual cortical processing. *NeuroImage*, *244*, 118613.
- 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., Smith, F. W., Schyns, P. G., & Muckli, L. (2013). Decoding face categories in diagnostic subregions of primary visual cortex. *Eur. J. Neurosci.*, *37*(7), 1130–1139.
- Petro, L. S., Vizioli, L., & Muckli, L. (2014). Contributions of cortical feedback to sensory processing in primary visual cortex. *Frontiers in psychology*, *5*, 1223.
- Peyrin, C., Mermillod, M., Chokron, S., & Marendaz, C. (2006). Effect of temporal constraints on hemispheric asymmetries during spatial frequency processing. *Brain Cogn.*, *62*(3), 214–220.
- Peyrin, C., Michel, C. M., Schwartz, S., Thut, G., Seghier, M., Landis, T., Marendaz, C., & Vuilleumier, P. (2010). The neural substrates and timing of top-down processes during coarse-to-fine categorization of visual scenes: A combined fMRI and ERP study. *J. Cogn. Neurosci.*, *22*(12), 2768–2780.
- Pitcher, D., Dilks, D. D., Saxe, R. R., Triantafyllou, C., & Kanwisher, N. (2011). Differential selectivity for dynamic versus static information in face-selective cortical regions. *Neuroimage*, *56*(4), 2356–2363.
- Pitcher, D., Walsh, V., Yovel, G., & Duchaine, B. (2007). Tms evidence for the involvement of the right occipital face area in early face processing. *Current Biology*, *17*(18), 1568–1573.
- Polat, U., & Sagi, D. (1993). Lateral interactions between spatial channels: Suppression and facilitation revealed by lateral masking experiments. *Vision research*, *33*(7), 993–999.
- Polyak, S. L. (1941). The retina.
- Power, J. D., Mitra, A., Laumann, T. O., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2014). Methods to detect, characterize, and remove motion artifact in resting state fmri. *Neuroimage*, *84*, 320–341.
- Puce, A., Allison, T., Asgari, M., Gore, J. C., & McCarthy, G. (1996). Differential sensitivity of human visual cortex to faces, letterstrings, and textures: A functional magnetic resonance imaging study. *Journal of neuroscience*, *16*(16), 5205–5215.
- Puri, A. M., Wojciulik, E., & Ranganath, C. (2009). Category expectation modulates baseline and stimulus-evoked activity in human inferotemporal cortex. *Brain research*, *1301*, 89–99.
- Pyles, J. A., Verstynen, T. D., Schneider, W., & Tarr, M. J. (2013). Explicating the face perception network with white matter connectivity. *PloS one*, *8*(4), e61611.
- Quek, G. L., Liu-Shuang, J., Goffaux, V., & Rossion, B. (2018). Ultra-coarse, single-glance human face detection in a dynamic visual stream. *NeuroImage*, *176*, 465–476.

- Rademaker, R. L., Chunharas, C., & Serences, J. T. (2019). Coexisting representations of sensory and mnemonic information in human visual cortex. *Nat. Neurosci.*, *22*(8), 1336–1344.
- Rajaei, K., Mohsenzadeh, Y., Ebrahimpour, R., & Khaligh-Razavi, S.-M. (2019). Beyond core object recognition: Recurrent processes account for object recognition under occlusion. *PLoS computational biology*, *15*(5), e1007001.
- Rao, R. P., & 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.
- Ress, D., & Heeger, D. J. (2003). Neuronal correlates of perception in early visual cortex. *Nature neuroscience*, *6*(4), 414–420.
- Retter, T. L., Jiang, F., Webster, M. A., & Rossion, B. (2020). All-or-none face categorization in the human brain. *NeuroImage*, *213*, 116685.
- Rieger, J. W., Gegenfurtner, K. R., Schalk, F., Koechy, N., Heinze, H.-J., & Grueschow, M. (2013). Bold responses in human v1 to local structure in natural scenes: Implications for theories of visual coding. *Journal of vision*, *13*(2), 19–19.
- Rieke, F., Warland, D., van Steveninck, R. d. R., & Bialek, W. (1996). Spikes: Exploring the neural code, a bradford book.
- Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature neuroscience*, *2*(11), 1019–1025.
- Riesenhuber, M., & Poggio, T. (2000). Models of object recognition. *Nature neuroscience*, *3*(11), 1199–1204.
- Ro, T., Breitmeyer, B., Burton, P., Singhal, N. S., & Lane, D. (2003). Feedback contributions to visual awareness in human occipital cortex. *Curr. Biol.*, *13*(12), 1038–1041.
- Rockland, K. S., & Lund, J. S. (1982). Widespread periodic intrinsic connections in the tree shrew visual cortex. *Science*, *215*(4539), 1532–1534.
- Rockland, K. S., & Lund, J. S. (1983). Intrinsic laminar lattice connections in primate visual cortex. *Journal of Comparative Neurology*, *216*(3), 303–318.
- Rockland, K. S., & Pandya, D. N. (1979). Laminar origins and terminations of cortical connections of the occipital lobe in the rhesus monkey. *Brain research*, *179*(1), 3–20.
- 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., & Spekreijse, H. (1998). Object-based attention in the primary visual cortex of the macaque monkey. *Nature*, *395*(6700), 376–381.
- Roland, P. E. (2010). Six principles of visual cortical dynamics. *Front. Syst. Neurosci.*, *4*, 28.

- Roland, P. E., Hanazawa, A., Undeman, C., Eriksson, D., Tompa, T., Nakamura, H., Valentinienė, S., & Ahmed, B. (2006). Cortical feedback depolarization waves: A mechanism of top-down influence on early visual areas. *Proc. Natl. Acad. Sci. U. S. A.*, *103*(33), 12586–12591.
- Rolls, E. T., & Tovee, M. J. (1994). Processing speed in the cerebral cortex and the neurophysiology of visual masking. *Proc. Biol. Sci.*, *257*(1348), 9–15.
- Rolls, E. T., Tovee, M. J., & Panzeri, S. (1999). The neurophysiology of backward visual masking: Information analysis. *J. Cogn. Neurosci.*, *11*(3), 300–311.
- Rosenblum, L. D., Yakel, D. A., Baseer, N., Panchal, A., Nodarse, B. C., & Niehus, R. P. (2002). Visual speech information for face recognition. *Perception & Psychophysics*, *64*, 220–229.
- Rossi, A. F., Desimone, R., & Ungerleider, L. G. (2001). Contextual modulation in primary visual cortex of macaques. *Journal of Neuroscience*, *21*(5), 1698–1709.
- Rossion, B., Dricot, L., Devolder, A., Bodart, J. M., Crommelinck, M., De Gelder, B., & Zoontjes, R. (2000). Hemispheric asymmetries for whole-based and part-based face processing in the human fusiform gyrus. *J. Cogn. Neurosci.*, *12*(5), 793–802.
- Rossion, B. (2009). Clarifying the functional neuro-anatomy of face perception by single case neuroimaging studies of acquired prosopagnosia. *Cortical Mechanisms of Vision*, 171–207.
- Rossion, B., et al. (2014). Understanding face perception by means of prosopagnosia and neuroimaging. *Frontiers in bioscience*, *6*(258), e307.
- Rossion, B., Caldara, R., Seghier, M., Schuller, A.-M., 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 Cogn.*, *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.
- Rossion, B., Schiltz, C., Robaye, L., Pirenne, D., & Crommelinck, M. (2001). How does the brain discriminate familiar and unfamiliar faces?: A pet study of face categorical perception. *Journal of cognitive neuroscience*, *13*(7), 1019–1034.
- Rotshtein, P., Vuilleumier, P., Winston, J., Driver, J., & Dolan, R. (2007). Distinct and convergent visual processing of high and low spatial frequency information in faces. *Cerebral Cortex*, *17*(11), 2713–2724.
- Rubin, G., & Siegel, K. (1984). Recognition of low-pass filtered faces and letters. *Invest. Ophthalmol. Vis. Sci. Suppl.*, *25*, 71.
- Rust, N. C., & DiCarlo, J. J. (2010). Selectivity and tolerance (invariance) both increase as visual information propagates from cortical area v4 to it. *Journal of Neuroscience*, *30*(39), 12978–12995.

- Ryan, T. A., & Schwartz, C. B. (1956). Speed of perception as a function of mode of representation. *The American journal of psychology*, 69(1), 60–69.
- Satterthwaite, F. E. (1941). Synthesis of variance. *Psychometrika*, 6(5), 309–316.
- Schiltz, C., & Rossion, B. (2006). Faces are represented holistically in the human occipito-temporal cortex. *Neuroimage*, 32(3), 1385–1394.
- Schmolesky, M. T., Wang, Y., Hanes, D. P., Thompson, K. G., Leutgeb, S., Schall, J. D., & Leventhal, A. G. (1998). Signal timing across the macaque visual system. *Journal of neurophysiology*, 79(6), 3272–3278.
- Scholte, H. S., Jolij, J., Fahrenfort, J. J., & Lamme, V. A. F. (2008). Feedforward and recurrent processing in scene segmentation: Electroencephalography and functional magnetic resonance imaging. *J. Cogn. Neurosci.*, 20(11), 2097–2109.
- Scholte, H. S., Jolij, J., & Lamme, V. (2006). The cortical processing dynamics of edge detection and scene segmentation. *The first half second*, 73–88.
- Schuurmans, J. P., Bennett, M. A., Petras, K., & Goffaux, V. (2023). Backward masking reveals coarse-to-fine dynamics in human v1. *NeuroImage*, 274, 120139.
- 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.
- Schyns, P. G. (1998). Diagnostic recognition: Task constraints, object information, and their interactions. *Cognition*, 67(1-2), 147–179.
- Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time- and spatial-scale-dependent scene recognition. *Psychol. Sci.*, 5(4), 195–200.
- Sclar, G., Maunsell, J. H., & Lennie, P. (1990). Coding of image contrast in central visual pathways of the macaque monkey. *Vision research*, 30(1), 1–10.
- Scouten, A., Papademetris, X., & Constable, R. T. (2006). Spatial resolution, signal-to-noise ratio, and smoothing in multi-subject functional mri studies. *Neuroimage*, 30(3), 787–793.
- Seijdel, N., Tsakmakidis, N., De Haan, E. H., Bohte, S. M., & Scholte, H. S. (2020). Depth in convolutional neural networks solves scene segmentation. *PLoS computational biology*, 16(7), e1008022.
- Self, M. W., van Kerkoerle, T., Super, H., & Roelfsema, P. R. (2013). Distinct roles of the cortical layers of area v1 in figure-ground segregation. *Current biology*, 23(21), 2121–2129.
- Sergent, J., Ohta, S., & Macdonald, B. (1992). Functional neuroanatomy of face and object processing: A positron emission tomography study. *Brain*, 115(1), 15–36.
- Sergent, J., & Poncet, M. (1990). From covert to overt recognition of faces in a prosopagnosic patient. *Brain*, 113(4), 989–1004.
- Sergerie, K., Lepage, M., & Armony, J. L. (2005). A face to remember: Emotional expression modulates prefrontal activity during memory formation. *Neuroimage*, 24(2), 580–585.

- Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proceedings of the national academy of sciences*, 104(15), 6424–6429.
- Shoham, D., Hübener, M., Schulze, S., Grinvald, A., & Bonhoeffer, T. (1997). Spatio-temporal frequency domains and their relation to cytochrome oxidase staining in cat visual cortex. *Nature*, 385(6616), 529–533.
- Simoncelli, E. P. (2003). Vision and the statistics of the visual environment. *Current opinion in neurobiology*, 13(2), 144–149.
- Sinha, P. (2002). Identifying perceptually significant features for recognizing faces. *Human Vision and Electronic Imaging VII*, 4662, 12–21.
- Smith, F. W., & Muckli, L. (2010). Nonstimulated early visual areas carry information about surrounding context. *Proc. Natl. Acad. Sci. U. S. A.*, 107(46), 20099–20103.
- Smith, M. L., Gosselin, F., & Schyns, P. G. (2012). Measuring internal representations from behavioral and brain data. *Current Biology*, 22(3), 191–196.
- Smith, S. A., Beaulieu, J. M., & Donoghue, M. J. (2010). An uncorrelated relaxed-clock analysis suggests an earlier origin for flowering plants. *Proceedings of the National Academy of Sciences*, 107(13), 5897–5902.
- Snowden, R. J., Treue, S., & Andersen, R. A. (1992). The response of neurons in areas v1 and mt of the alert rhesus monkey to moving random dot patterns. *Experimental Brain Research*, 88, 389–400.
- Snyder, L. H., Batista, A., & Andersen, R. A. (1997). Coding of intention in the posterior parietal cortex. *Nature*, 386(6621), 167–170.
- Somers, D. C., Nelson, S. B., & Sur, M. (1995). An emergent model of orientation selectivity in cat visual cortical simple cells. *Journal of neuroscience*, 15(8), 5448–5465.
- Sorger, B., Goebel, R., Schiltz, C., & Rossion, B. (2007). Understanding the functional neuroanatomy of acquired prosopagnosia. *Neuroimage*, 35(2), 836–852.
- Spoerer, C. J., McClure, P., & Kriegeskorte, N. (2018). Corrigendum: Recurrent convolutional neural networks: A better model of biological object recognition. *Frontiers in Psychology*, 9, 1695.
- Steeves, J. K., Culham, J. C., Duchaine, B. C., Pratesi, C. C., Valyear, K. F., Schindler, I., Humphrey, G. K., Milner, A. D., & Goodale, M. A. (2006). The fusiform face area is not sufficient for face recognition: Evidence from a patient with dense prosopagnosia and no occipital face area. *Neuropsychologia*, 44(4), 594–609.
- Stelzer, J., Lohmann, G., Mueller, K., Buschmann, T., & Turner, R. (2014). Deficient approaches to human neuroimaging. *Frontiers in Human Neuroscience*, 8, 462.
- Stigliani, A., Weiner, K. S., & Grill-Spector, K. (2015). Temporal processing capacity in high-level visual cortex is domain specific. *Journal of Neuroscience*, 35(36), 12412–12424.

- Stromeyer, C. F., 3rd, & Julesz, B. (1972). Spatial-frequency masking in vision: Critical bands and spread of masking. *J. Opt. Soc. Am.*, 62(10), 1221–1232.
- Sugiura, M., Sassa, Y., Jeong, H., Horie, K., Sato, S., & Kawashima, R. (2008). Face-specific and domain-general characteristics of cortical responses during self-recognition. *Neuroimage*, 42(1), 414–422.
- Summerfield, C., Egner, T., Mangels, J., & Hirsch, J. (2006). Mistaking a house for a face: Neural correlates of misperception in healthy humans. *Cerebral cortex*, 16(4), 500–508.
- Super, H., Spekreijse, H., & Lamme, V. A. (2001). A neural correlate of working memory in the monkey primary visual cortex. *Science*, 293(5527), 120–124.
- Supèr, H., Spekreijse, H., & Lamme, V. A. (2001). Two distinct modes of sensory processing observed in monkey primary visual cortex (v1). *Nature neuroscience*, 4(3), 304–310.
- Taira, M., Mine, S., Georgopoulos, A., Murata, A., & Sakata, H. (1990). Parietal cortex neurons of the monkey related to the visual guidance of hand movement. *Experimental brain research*, 83, 29–36.
- Taira, M., Tsutsui, K.-I., Jiang, M., Yara, K., & Sakata, H. (2000). Parietal neurons represent surface orientation from the gradient of binocular disparity. *Journal of neurophysiology*, 83(5), 3140–3146.
- Takagaki, K., Zhang, C., Wu, J.-Y., & Lippert, M. T. (2008). Crossmodal propagation of sensory-evoked and spontaneous activity in the rat neocortex. *Neurosci. Lett.*, 431(3), 191–196.
- Talebi, V., & Baker, C. L. (2012). Natural versus synthetic stimuli for estimating receptive field models: A comparison of predictive robustness. *Journal of neuroscience*, 32(5), 1560–1576.
- Tanaka, H., & Sawada, R. (2022). Dynamics and mechanisms of contrast-dependent modulation of spatial-frequency tuning in the early visual cortex. *J. Neurosci.*, 42(37), 7047–7059.
- Tang, H., Schrimpf, M., Lotter, W., Moerman, C., Paredes, A., Ortega Caro, J., Hardesty, W., Cox, D., & Kreiman, G. (2018). Recurrent computations for visual pattern completion. *Proceedings of the National Academy of Sciences*, 115(35), 8835–8840.
- Tanskanen, T., Näsänen, R., Ojanpää, H., & Hari, R. (2007). Face recognition and cortical responses: Effect of stimulus duration. *Neuroimage*, 35(4), 1636–1644.
- Thome, I., Hohmann, D. M., Zimmermann, K. M., Smith, M. L., Kessler, R., & Jansen, A. (2021). I spy with my little eye, something that is a face. . “: A brain network for illusory face detection. *Cereb Cortex*, 32(1), 137–157.
- Thompson, K. G., & Schall, J. D. (1999). The detection of visual signals by macaque frontal eye field during masking. *Nature neuroscience*, 2(3), 283–288.
- Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *nature*, 381(6582), 520–522.

- Tolhurst, D. (1973). Separate channels for the analysis of the shape and the movement of a moving visual stimulus. *The Journal of Physiology*, 231(3), 385.
- Tolhurst, D., & Thompson, I. (1982). Organization of neurones preferring similar spatial frequencies in cat striate cortex. *Experimental Brain Research*, 48(2), 217–227.
- Tomita, H., Ohbayashi, M., Nakahara, K., Hasegawa, I., & Miyashita, Y. (1999). Top-down signal from prefrontal cortex in executive control of memory retrieval. *Nature*, 401(6754), 699–703.
- Tootell, R. B., Reppas, J. B., Kwong, K. K., Malach, R., Born, R. T., Brady, T. J., Rosen, B. R., & Belliveau, J. W. (1995). Functional analysis of human mt and related visual cortical areas using magnetic resonance imaging. *Journal of Neuroscience*, 15(4), 3215–3230.
- Tootell, R. B., Silverman, M. S., & De Valois, R. L. (1981). Spatial frequency columns in primary visual cortex. *Science*, 214(4522), 813–815.
- Torralba, A., & Oliva, A. (2003). Statistics of natural image categories. *Network: computation in neural systems*, 14(3), 391.
- Tsao, D. Y., Freiwald, W. A., Tootell, R. B., & Livingstone, M. S. (2006). A cortical region consisting entirely of face-selective cells. *Science*, 311(5761), 670–674.
- Tsao, D. Y., & Livingstone, M. S. (2008). Mechanisms of face perception. *Annu. Rev. Neurosci.*, 31, 411–437.
- Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., & Gee, J. C. (2010). N4ITK: Improved N3 bias correction. *IEEE Trans. Med. Imaging*, 29(6), 1310–1320.
- Ungerleider, L. G., & Haxby, J. V. (1994). whatand wherein the human brain. *Current opinion in neurobiology*, 4(2), 157–165.
- Van Der Zwaag, W., Marques, J., Kober, T., Glover, G., Gruetter, R., & Krueger, G. (2012). Temporal snr characteristics in segmented 3d-epi at 7t. *Magnetic resonance in medicine*, 67(2), 344–352.
- Van Dromme, I. C., Premereur, E., Verhoef, B.-E., Vanduffel, W., & Janssen, P. (2016). Posterior parietal cortex drives inferotemporal activations during three-dimensional object vision. *PLoS Biol.*, 14(4), e1002445.
- 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.
- Visconti di Oleggio Castello, M., & Gobbini, M. I. (2015). Familiar face detection in 180ms. *PloS one*, 10(8), e0136548.
- Vizioli, L., Steen, M., Logan, D., Mehmet, A., Federico, D. M., Essa, Y., & Kamil, U. (2020). A paradigm change in functional brain mapping: Suppressing the thermal noise in fmri. *BioRxiv*, 2020–11.
- Vlamings, P. H., Goffaux, V., & 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*, 9(5), 12–12.

- Wang, H., Lian, Y., Wang, A., Chen, E., & Liu, C. (2023). Face motion form at learning influences the time course of face spatial frequency processing during test. *Biological Psychology*, 183, 108691.
- Wang, T., Li, Y., Yang, G., Dai, W., Yang, Y., Han, C., Wang, X., Zhang, Y., & Xing, D. (2020). Laminar subnetworks of response suppression in macaque primary visual cortex. *Journal of Neuroscience*, 40(39), 7436–7450.
- Warrington, E. K., & James, M. (1967). An experimental investigation of facial recognition in patients with unilateral cerebral lesions. *Cortex*, 3(3), 317–326.
- Watt, R. (1987). Scanning from coarse to fine spatial scales in the human visual system after the onset of a stimulus. *JOSA A*, 4(10), 2006–2021.
- Weiner, K. S., & Grill-Spector, K. (2013). Neural representations of faces and limbs neighbor in human high-level visual cortex: Evidence for a new organization principle. *Psychol. Res.*, 77(1), 74–97.
- Westheimer, G. (2001). The fourier theory of vision. *Perception*, 30(5), 531–541.
- Wibral, M., Bledowski, C., Kohler, A., Singer, W., & Muckli, L. (2009). The timing of feedback to early visual cortex in the perception of long-range apparent motion. *Cereb. Cortex*, 19(7), 1567–1582.
- Williford, T., & Maunsell, J. H. (2006). Effects of spatial attention on contrast response functions in macaque area v4. *Journal of neurophysiology*, 96(1), 40–54.
- Worgotter, F., & Koch, C. (1991). A detailed model of the primary visual pathway in the cat: Comparison of afferent excitatory and intracortical inhibitory connection schemes for orientation selectivity. *Journal of Neuroscience*, 11(7), 1959–1979.
- Wyatte, D., Curran, T., & O'Reilly, R. (2012). The limits of feedforward vision: Recurrent processing promotes robust object recognition when objects are degraded. *Journal of Cognitive Neuroscience*, 24(11), 2248–2261.
- Wyatte, D., Jilk, D. J., & O'Reilly, R. C. (2014). Early recurrent feedback facilitates visual object recognition under challenging conditions. *Front. Psychol.*, 5, 674.
- Xu, X., Anderson, T. J., & Casagrande, V. A. (2007). How do functional maps in primary visual cortex vary with eccentricity? *Journal of Comparative Neurology*, 501(5), 741–755.
- Yan, X., Goffaux, V., & Rossion, B. (2022). Coarse-to-fine (r) automatic familiar face recognition in the human brain. *Cerebral Cortex*, 32(8), 1560–1573.
- Yan, X., & Rossion, B. (2020). A robust neural familiar face recognition response in a dynamic (periodic) stream of unfamiliar faces. *cortex*, 132, 281–295.
- Yang, H., Susilo, T., & Duchaine, B. (2016). The anterior temporal face area contains invariant representations of face identity that can persist despite the loss of right ffa and ofa. *Cerebral Cortex*, 26(3), 1096–1107.
- Yip, A. W., & Sinha, P. (2002). Contribution of color to face recognition. *Perception*, 31(8), 995–1003.

- Yoshor, D., Bosking, W. H., Ghose, G. M., & Maunsell, J. H. (2007). Receptive fields in human visual cortex mapped with surface electrodes. *Cerebral cortex*, 17(10), 2293–2302.
- Young, A. W., Newcombe, F., Haan, E. H. d., Small, M., & Hay, D. C. (1993). Face perception after brain injury: Selective impairments affecting identity and expression. *Brain*, 116(4), 941–959.
- Yue, X., Nasr, S., Devaney, K. J., Holt, D. J., & Tootell, R. B. (2013). Fmri analysis of contrast polarity in face-selective cortex in humans and monkeys. *Neuroimage*, 76, 57–69.
- Yue, X., Tjan, B. S., & Biederman, I. (2006). What makes faces special? *Vision research*, 46(22), 3802–3811.
- Zachariou, V., Nikas, C. V., Safiullah, Z. N., Gotts, S. J., & Ungerleider, L. G. (2017). Spatial mechanisms within the dorsal visual pathway contribute to the configural processing of faces. *Cereb. Cortex*, 27(8), 4124–4138.
- Zeki, S. M. (1974). Functional organization of a visual area in the posterior bank of the superior temporal sulcus of the rhesus monkey. *The Journal of physiology*, 236(3), 549–573.
- Zhan, M., Pallier, C., Agrawal, A., Dehaene, S., & Cohen, L. (2023). Does the visual word form area split in bilingual readers? a millimeter-scale 7-t fmri study. *Science Advances*, 9(14), eadf6140.
- Zhang, H., Liu, J., Huber, D. E., Rieth, C. A., Tian, J., & Lee, K. (2008). Detecting faces in pure noise images: A functional mri study on top-down perception. *Neuroreport*, 19(2), 229–233.
- Zhang, Y., Brady, M., & Smith, S. (2001). Segmentation of brain MR images through a hidden markov random field model and the expectation-maximization algorithm. *IEEE Trans. Med. Imaging*, 20(1), 45–57.
- Zhao, L., & Chubb, C. (2001). The size-tuning of the face-distortion after-effect. *Vision research*, 41(23), 2979–2994.
- Zhao, M., Hayward, W. G., & Bülthoff, I. (2014). Face format at encoding affects the other-race effect in face memory. *Journal of Vision*, 14(9), 6–6.
- Zhen, Z., Yang, Z., Huang, L., Kong, X.-z., Wang, X., Dang, X., Huang, Y., Song, Y., & Liu, J. (2015). Quantifying interindividual variability and asymmetry of face-selective regions: A probabilistic functional atlas. *Neuroimage*, 113, 13–25.
- Zipser, K., Lamme, V. A., & Schiller, P. H. (1996). Contextual modulation in primary visual cortex. *Journal of Neuroscience*, 16(22), 7376–7389.