1. General Introduction

Spatial perception concerns our ability to sense the position, movement, orientation – and to some extend the size and the shape – of objects and of ourselves in relation to our environment. It is a crucial part of human visual perception since we use spatial information for many crucial tasks, e.g., to localize ourselves in our environment, to direct our gaze and attention, and to reach out to and grasp objects. Our efficiency in determining a visual object’s position in space is affected by various factors. Complex and crowded scenes, movement of objects, but also body, head and eye movements can cause inaccuracies in determining the perceived position of an object in space, its position relative to our body or its position relative to other objects.

The perceived position – and the involved information processing by our visual system – can be measured by using psychophysical methods and eye-tracking. In this thesis, I have used these techniques to study spatial perception in order to better understand how human observers localize and identify objects. In addition, I have used neuroimaging to investigate the neurobiological basis of spatial perception. For this, I have designed and used various non-conventional stimuli to examine where and how the brain processes visual spatial information.

The overall aim of my thesis is to better understand the role of a number of factors such as physical object location, object identity, orientation and contour integration in human spatial perception. The next section will detail this aim by outlining the scope of each experimental chapter. While I briefly describe the results that I found, these are described and discussed in more detail in the General Discussion (chapter 7). In proceeding sections, I will inform on the biological bases, the visual phenomena, and the methods that are described and used in this thesis.

1.1. Outline

One of the most easily observed characteristics of natural human visual behavior is the frequent shifting of our eye movements (about 3–4 times per second). We make such shifts, called saccades, because only the central part of our visual field can be used to analyze visual objects in detail. In between saccades, our eyes remain at a more or less fixed position during which the visual system extracts information from the environment for perception and for planning a saccade to the next relevant location. Despite the relevance of eye movements for natural human visual behavior, many fields of vision science have tended to ignore eye movements. In others, the question has arisen whether preparing for, or responding by means of, eye movements would reveal
the same kind of processing as revealed when observers are required to fixate and respond by hand. Recent studies have suggested that spatial perception might be more accurate if observers can use eye-movements (Deubel, 2008; Harrison, Mattingley, & Remington, 2013). However, the perceived position of an object is not always consistent for all tasks. That means, depending on the task, the same physical position might result in a different perceived position. This is so because different tasks (e.g. manual pointing, making an eye movement to an object, verbal report) might emphasize different types of position information such as directionality, absolute position or depth (Smeets & Brenner, 1995).

Chapter 2: In this chapter, I studied the question to what extend the eyes follow the perceived position of a visual object. This is an often discussed and studied question, but in many studies widely different eye movement and perception tasks have been compared. In my study, this question is examined by using very similar and simple localization tasks in both cases. My main finding of this chapter is that eye movements and visual-decision-making are based on shared neural representations.

Chapter 3: Further studies have suggested that eye movements would also affect object recognition. This would affect the recognition of visual objects in more complex environments, when they are affected by "crowding". In the visual crowding phenomenon, recognition performance is declined when a target is surrounded by similar objects. However, different studies showed contradictory findings on this topic, with one showing an improvement (Harrison, Mattingley, et al., 2013) and another a decline (Harrison, Retell, Remington, & Mattingley, 2013) in recognition performance. In a behavioral crowding experiment, I tried to answer whether a crowded object is differently perceived when observers have to select a target object by making an eye movement to it, compared to a condition in which they have to fixate their eyes and respond via a button press. To properly address this issue – and unlike many previous studies – I minimized the differences between the tasks used. I found that crowding strength (recognition performance for the features of an object) is preserved when responding by eye. Therefore, I found no evidence that (preparing to make) eye-movements affect recognition performance. The results of this chapter also made it possible to study our hypothesis of chapter 4 by means of measuring eye movements.

Chapter 4: In this chapter, I ask whether the complexity of the visual environment similarly affects the accuracy of perceived position and identity information. For this, I relied on the "crowding" phenomenon again. A lot is already known about how crowding affects identity perception (Pelli, 2008; van den Berg, Johnson, Anton, Schepers, & Cornelissen, 2012). The identity of the objects in a scene could also affect where a crowded target object is perceived. Since the similarity level between the target and the surrounding objects influences the crowding magnitude, I was curious to verify if the perceived position of a crowded object was similarly affected. By studying the influence of crowding on the perceived position of objects, my aim was to gain further insight in the sources of ambiguity underlying location perception in the human visual system. I found that location and identity perception are similarly affected by the strength of crowding.

Chapter 5: Visual cortex is divided into different areas, –with each area containing a representation of a part of the visual field. Typically, a luminance-contrast (e.g.
A (General Discussion).

1.2. Background

This section will inform on the biological bases, the visual phenomena, and the methods that are described and used in this thesis.

1.2.1. Visual Pathways

The visual system makes use of the information contained in the light reaching our eyes in order to interpret and make sense of the environment. The visual perceptual process starts when an image is projected via the eye lens onto the layer of retinal photoreceptors at the back of the eye. The retina contains two types of receptor, the rods and the cones. Rods are responsible for vision at low light levels while cones mediate vision during medium to high light levels. There are three types of cones that are sensitive to short, middle and high wavelengths. They mediate the signals for color vision. The central part of the retina is called the fovea and contains the highest density of cones, thereby mediating detailed vision. The photoreceptors, via the bipolar cells, transmit signals to the ganglion cells. Through the optic nerves, which cross at the optic chiasm, the signals of the ganglion cells reach the lateral geniculate nucleus in the thalamus as well as a number of other subcortical nuclei. After the thalamus, the optic...
radiations project onto the visual cortex, which is subdivided in a number of distinct functional areas (Figure 1).

![Diagram of the visual cortex](image)

**Figure 1:** An illustration depicting the visual pathway and the visual system (Logothetis, 1999).

### 1.2.2. Visual Cortex

Primary visual cortex, or V1, has a retinotopic organization, which means that adjacent points in the visual field are processed by adjacent points in the cortex. This retinotopic organization indicates that there is a systematic one-to-one projection from the retina to the cortical area. By presenting a controlled visual stimulus, systematic visual field maps can be obtained. This is called retinotopic mapping and is discussed further below. Until now, nearly 20 visual areas have been mapped (Figure 2). Each of these regions receives information either from half of the visual field (e.g. V1) or of a quarter visual field when information from the upper and lower visual fields is divided over dorsal and ventral sections (e.g. the case for areas V2 and V3). In the higher visual areas, neurons have complex tuning functions. For example, in the fusiform face area (FFA), a neuron may fire when an object that matches its filter function (e.g. a familiar face) appears in its receptive field (Gauthier, Skudlarski, Gore, & Anderson, 2000).

### 1.2.3. Spatial Information Processing

Localizing an object is a complicated process and requires multiple stages of processing (figure 3). Knowing an object’s position by its retinal position is not enough to act upon
it (such as making an eye movement (saccade) to it). To do so, our visual system needs to calculate where an object is relative to our eye, our body or our hands. Orientation of the eyes needs to be taken into account for this. This requires transferring information between different coordinate systems, to relate the retinal image and the posture of the body parts that will be used, e.g. for grasping. Localization of objects is studied in chapter 2.

![Visual field maps in the human visual system](Wandell, Dumoulin, & Brewer, 2007). The (+) and (-) signs correspond to the upper and lower part of the visual field and indicate where these are processed in the brain respectively. The circle indicates the foveal representation. Lateral occipital cortex (LO-x), ventral occipital cortex (VO-x) and intraparietal sulcus (IPS-x) are shown with their abbreviations.

There are additional influences on localization by population coding and lateral interactions between neurons. These effects are dependent on the number of elements or objects in the visual field. This is so because when there is an increased number of
objects in the visual field, the receptive fields\(^1\) that encode the different objects may overlap, causing interactions in the population encoding. This could be a restrictive mechanism, as it limits or mixes the information about distinct objects in the visual field (see also crowding below). Another mechanism is lateral inhibition. Neurons do not respond individually to a certain stimulus but form a network with other neurons that selectively respond to particular orientations and spatial frequencies. Neurons make lateral inhibitory connections with neurons that respond to – usually similar – objects or features in nearby locations (Polat & Sagi, 1993). It is not obvious that a neuroimaging method (such as fMRI) can be used to resolve neuronal activity at a sufficient level of detail and thus detect such lateral interactions. The study in chapter five uses orientation contrast to map out visual areas and to determine cortical population properties. It presumably relies on inhibitory interactions between neurons to obtain differential fMRI responses.

Figure 3: Zollner illusion. The long black lines are seen unparalleled even though in reality they are parallel. The orientation of the inducing lines (short lines) is causing the illusory orientation of the induced line (test line). This is an example of a shared feature (orientation) of a group of objects modifying the perceived relative position to other objects (Kitaoka & Ishihara, 2000).

1.2.4. Crowding, grouping and contour integration

Crowding is a breakdown of object recognition when the target object is surrounded by other similar objects (that are referred to as flankers). In a typical crowding task, observers fixate on a point while a target is presented in the periphery surrounded by flankers and they carry out a task by reporting an aspect of the target object. According to Bouma’s law (Bouma, 1970), flankers should be at least within half the proximity of the distance from target to fixation to the target object for crowding to take place. Similarity of objects is also known to increase crowding magnitude. Various aspects of the target such as perceived location, orientation, color and identity are affected by crowding (see figure 4 for an example of crowding).

\(^1\) The receptive field of a sensory neuron is the spatial region in the visual field in which a stimulation could trigger the firing of that neuron.
There exist different theories to explain and model the mechanisms that are involved in crowding. According to the substitution/confusion theory, the identity of objects are confused with each other implying that our visual system can identify a crowded object accurately but might be primarily confused about its location in space (Huckauf & Heller, 2002; Strasburger, 2005). According to the feature integration or pooling theory, the visual system might average the visual aspects (features) of nearby objects and thereby assign an inaccurate identity to the resulting jumbled percept (Dakin, Cass, Greenwood, & Bex, 2010; Greenwood, Bex, & Dakin, 2009; Parkes, Lund, Angelucci, Solomon, & Morgan, 2001; van den Berg, Roerdink, & Cornelissen, 2010). Crowding is studied in chapter 4.

Figure 4: An example of crowding. When fixating on the red point on the left, it is fairly easy to identify the scissor on the bottom image while it is more challenging to say on the top (Chaney, Fischer, & Whitney, 2014).

The visual system assembles objects into groups or continuous structures (contours) according to different factors such as similarity of color, proximity, size, orientation and common fate in moving objects (Brooks, 2014). This grouping phenomenon is exploited in chapter five to create a new type of retinotopy stimulus based on orientation information. Contour Integration can be considered a grouping process whereby continuity is perceived between aligned elements (figure 5). The role of contour integration in object perception is examined in chapter 6.
Figure 5: Contour integration. Continuously oriented Gabor signals are relatively easily distinguished from a random distribution of such signals. Example of (a) aligned gabor patches as a semicircular curve presented with a background of randomly oriented gabor patches (b) gabor patches pop out even more when the continuation is arranged as a circle. (c) stimulus in b shown with an overlaid contour (adapted from Kovács & Julesz, 1993).

1.3. Methods

In this thesis I used manual responses, eye movement recordings, and functional MRI. They are all non-invasive techniques that can be used to study human visual behavior and the underlying neural organization.

1.3.1. Manual Responses

Manual responses are the most common method used in psychophysics. Behavioral responses are often used to determine discrimination, detection or identification thresholds or to find biases in perception. There are various methods to determine discrimination thresholds of which I have used the two-alternative forced choice (2AFC) task which requires the participant to indicate which of two options was the perceived one (in which a decision is made e.g. according to the orientation, the location, or the shape of the object). The response frequencies for different stimuli or stimulus properties (e.g. orientation) can be used to construct a psychometric curve. This curve, in turn, can be used to determine a discrimination threshold. This is done by determining the stimulus magnitude required to reach a certain performance. I used this method in chapters 3 and 4 to determine the sensitivity of each individual observer. Using this information, I could adjust the properties of the experimental stimuli, such that they became equally discriminable for each observer. In general, using such individually tuned properties increases the sensitivity of the experimental procedure.

1.3.2. Eye Tracking

The visual system uses the eyes to collect information from the surroundings. We move our eyes for scrutinizing the parts of the visual field that we think may contain important information. We direct our gaze to an object to obtain more detailed information from it. When a scene is presented we use various eye movement strategies to retrieve the information necessary to carry out a given task. Eye movements can be made voluntarily or through reflexes. Typically, the voluntary motions are either
through making small jumps in horizontal or vertical direction (saccades) or tracking of a moving object (smooth pursuit). If the gaze remains in a stable position for at least 100-2000 ms it is considered a fixation. Measuring these fixations’ location, their duration and the order of fixations can give clues about the saliency of the information presented, and about how much information is retrieved or the time it takes to process certain aspects. Saccade and fixation metrics can be used to understand the cognitive processes used by the participants in order to perform a given task. Saccades can also be used to measure how accurately the human visual system is able to localize an object. In chapters 2-4 I use this technique to study a number of perceptual processes such as localization accuracy, target discrimination performance and response time.

1.3.3. Functional magnetic resonance imaging (fMRI)

When neurons become active, they use oxygen which is supplied by the blood. When a cortical region becomes active the amount of oxygenated and deoxygenated hemoglobin changes. The brain reacts to this by sending extra blood to the region. This is called the Blood Oxygenation Level Dependent (BOLD) hemodynamic response. As a consequence of this response, the magnetic properties of the blood change which can be picked up by functional magnetic resonance imaging (fMRI) (Ogawa & Lee, 1990). fMRI applies a high magnetic field to the brain tissue in order to align the proton spins and to capture the radiofrequency signal that is released when the spins relax again (Buxton & Diego, 2015). The speed of relaxation depends on the tissue and oxygenated/deoxygenated hemoglobin ratio. By frequently acquiring MRI signals, it is possible to monitor functional activations in the brain. This results in a time course of activations for all scanned locations in the brain. Typically, at a field strength of 3 Tesla, fMRI signals have a temporal resolution around 1-2 seconds and a spatial resolution of about 1-2 mm³.

1.3.4. fMRI signal analysis using traveling wave analyses

The traveling wave method (phase-encoded analysis; figure 6) can be used to analyze the activity in a voxel (a 3D volumetric pixel) over the time course of an experimental run (Engel et al., 1994). The stimulus typically used in this method is characterized by a certain on/off period. A sinusoidal signal at the stimulus alternation frequency is fitted to the voxel’s time course. Specifically for retinotopic mapping, the on/off cycles for each position in the visual field change systematically over time (hence the term "traveling" wave; figure 6). The phase of the fitted sinusoid can be used as an indication of the preferred position, while the amplitude of the fit indicates the response magnitude. I use this method in chapter 6 to determine the activity evoked by an illusory shape stimulus, that appears and disappears on the screen. However, in particular for retinotopic mapping, I have used a more recent approach (see the next section).
Figure 6: Two example stimuli frame from the expanding ring (A) and rotating wedge (B) stimulus. The stimulus moves in the direction of the arrows presented. These stimuli evoke a BOLD signal modulation between 1%–3%. (C) The time (phase) of the peak modulation varies smoothly across the cortical surface. Six stimulus repetitions are shown. In this case, space measures distance along the medial visual cortex (indicated by the dashed lines in panels D and E). The time delay (phase) defines the most effective stimulus eccentricity, derived by the ring stimuli (D) and angle derived by the wedge (E) in the stimulus sequence. Adapted from (Wandell, Dumoulin, & Brewer, 2007).

1.3.5. Retinotopic mapping using Population Receptive Field Modeling

The traveling wave method has been used to chart the visual field maps in the brain. Nevertheless, it is not based on a biologically plausible model since it relies on explaining the activity of each voxel as an individual unit and ignores the well-established principle of the receptive field. Moreover, because of excitatory and inhibitory lateral neuronal connections, neurons are sensitive to activity in neighboring
neurons. Together, the ensemble of the receptive fields of all neurons also create a receptive field, called the "population receptive field" (pRF). These aspects are taken into account in the so-called "pRF method" (figure 7). In its most basic form, this method models the pRF as a simple circular symmetric Gaussians that assigns a receptive field size (σ of the Gaussian) and a location in the visual field to each voxel. Next, it creates a predicted BOLD signal for each voxel as a combination of the model pRF, the stimulus information (the frame by frame stimulus over time) and convolves it with the hemodynamic response function (HRF). Next, it estimates the goodness of fit of this model to explain the observed BOLD activation according to each voxel's activity over time. Depending on the quality of the fit, the model is either adapted or accepted. This results in maps of the visual cortex in which all voxels have been assigned a location (eccentricity and polar angle in the visual field) as well as a receptive field size (an example can be seen in figure 3 of chapter 5). This method is used in chapters 5 and 6. In chapter 5, a new stimulus – based on orientation-contrast – is tested for its use for determining accurate maps of the visual cortex.

Figure 7: The pipeline of the population receptive field (pRF) analysis (Dumoulin & Wandell (2008)). For each voxel, the method models the pRF as a circular symmetric Gaussians that assigns a receptive field size (σ of the Gaussian) and a location (x, y) in the visual field. By convolving the pRF with the stimulus, a pRF activation is predicted. This predicted pRF activation is convolved with a hemodynamic response function (HRF) to estimate the predicted BOLD signal. This function is assumed to be the response of the system (as reflected by the MR signal) to aperiod of neural stimulation. By determining the error between the prediction and the data, the best fitting model for a particular region can be estimated. Adapted from Dumoulin & Wandell (2008).
1.4. References


