Scuola Internazionale Superiore di Studi Avanzati – Trieste

Functional evidence of hierarchical object processing in rat lateral extrastriate cortex

Candidate:
Eis Annavini

Supervisor:
Prof. Davide Zoccolan

Thesis submitted for the degree of *PhilosophiæDoctor*
SISSA – Trieste, January 26, 2018

Black then white are all I see in my infancy.
Red and yellow then came to be, reaching out to me.
Lets me see.

(Tool, Lateralus)

In the 60s, Marvin Minsky assigned a couple of undergrads to spend the summer programming a computer to use a camera to identify objects in a scene. He figured they'd have the problem solved by the end of the summer. Half a century later, we're still working on it.

(Randall Munroe, xkcd)

Contents

Acknowledgements

A scientific experiment, like any work of ingenuity, is rarely, if ever, accomplished alone. Thus, I would like to thank all the people without whom all of this never would have happened.

First and foremost my thanks goes to my supervisor Davide Zoccolan, for the guidance, the support, and all the insight he gave me. He really is the best supervisor I could ask for. A big thank you also goes to my family, without their constant love and support I literally wouldn't be here. Next, I extend my warmest thanks to my own well of wisdom, Federica Rosselli. Thanks for all the time you shared with me, and for teaching me "all I know on earth, and all I need to know" about surgeries. I would also like to thank Margherita Riggi, Rosilari Bellacosa, Giulio Matteucci, and Mattia D'Andola. Thank you for all the invaluable help you gave me in these four years, and for always believing in me, even when I didn't. Finally I would like to thank all my labmates and friends for the daily snippets of life we lived together. Rest assured I will cherish many of them for the rest of my life.

It's been one hell of a ride, and it's been worth every minute of it.

Abstract

Every day, the countless tasks we carry out during our life require us to answer a key question: "what is it I'm looking at?" Robustly recognize objects in spite of drastic changes in their appearance is a basic requirement of our survival, and a hallmark of the computational capacity of primate neocortex; our visual system is in fact able to solve this problem without any apparent effort and with surprising and unparalleled speed.

In order to make this possible the visual cortex must be able to build, at a certain point, a representation of a visual stimulus that is invariant to the so-called *identity preserving transformations* (i.e. rotations, translations, scaling, background and/or lighting variations). In primates it was shown that this task is accomplished through processing visual information along a dedicated cortical pathway called the ventral stream, originating in V1 and reaching its highest level in the inferotemporal cortex (DiCarlo, Zoccolan, et al., 2012; Rolls, 2012). This stream is made up, at least for what concerns core object recognition, by a series of feedforward modules organized in a hierarchical fashion. These areas sequentially build more and more complex representations, as their neuronal tuning properties change from a preference for very simple stimuli (e.g. oriented edges) in the low-level areas, to a selectivity for complex patterns in the inferotemporal cortex (Desimone et al., 1984; Kobatake and Tanaka, 1994).

We are, however, still far from figuring out how the ventral stream solves the invariance problem, as the monkey visual system is a dramatically complex machine, composed by more than 35 areas reciprocally connected by hundreds of pathways (Felleman and Van Essen, 1991). For this reason, and over growing concerns about the use of primates in research, rats have been considered with ever increasing interest as models of the mammalian visual system. Although far simpler and with a visual acuity ten times lower than primates, rats allow access to a wider array of experimental techniques (e.g. two photon imaging, calcium imaging, optogenetics), and recent studies have shown that they are able to invariantly recognize visual objects undergoing identity-preserving transformations (for a review, see Zoccolan, 2015). Anatomical (Coogan and Burkhalter, 1993; Sereno and Allman, 1991) and electrophysiological (Tafazoli, Safaai, et al., 2017) studies indicate that along the progression of rat extrastriate areas running lateral to primary visual cortex (V1) low-level information is gradually pruned in favor of representing more complex features (Tafazoli, Safaai, et al., 2017).

In this study we sought to further examine the validity of the rat as a

model of the mammalian visual system by investigating in deeper detail the representation of visual objects along the rat putative ventral stream, and how it varies across its different areas. To this extent we recorded in-vivo extracellular activity in anesthetized rats passively exposed to a rich stimulus set, designed to explore a wide range of visual features and transformations. We analyzed the response to the stimuli both at the single-unit level, using information theory to characterize the amount of information conveyed by neuron of different areas about features of different complexity, and at the population level, checking whether and how the complexity of the population code changes across areas and how the different features are encoded by the extrastriate areas.

Our analysis shows that the information about low-level properties of the stimuli is progressively reduced both at the single neuron and at the population level. Single neuron analyses show that the mutual information between neuronal response and low level stimulus properties (e.g. total luminosity or position) decreases along the putative ventral stream, while the relative weight of the information about mid- and high-level stimulus properties increased, all in accordance with the information conservation principle. At the population level we used two different clustering algorithms (complete linkage hierarchical clustering and K-means) to group the stimuli according to their representation and then compare the result with the properties of the stimulus set at different levels.

The results of the complete linkage hierarchical clustering show that in the early stages of visual processing stimuli sharing similar low-level properties get clustered together in the first nodes (the ones closer to the root) of the hierarchical tree, while the same does not happen in later stages. Then we divided the stimulus set in categories based on properties of different complexity and we compared these categories to the results of the K-means clustering in the different areas using Adjusted Mutual Information, a well-established clustering comparison metric. The comparison yields similar results to the single neuron analyses, with low level areas' populations carrying more information about low level stimulus properties and with information about high level properties, after filtering out the effect of the low-level ones, staying constant across the hierarchy and thus increasing in relative importance.

These results provide further evidence that the functional organization of rat lateral extrastriate areas resembles the one of the primate ventral stream, further promoting the viability of rodents as models of high-level visual processing in mammals.

Chapter 1

Introduction

1.1 The Core Recognition Problem

Picking the paper you were looking for from a cluttered desk, recognizing a friend in a crowded place or reading a PhD thesis are activities we perform with ease and that have great importance in our everyday lives. Humans and other nonhuman primates are able to recognize an object in spite of often dramatic changes in size, appearance, or lighting conditions, and do so extremely quickly, within a fraction of a second (DiCarlo, Zoccolan, et al., 2012). During behavioral tests, the reaction times reported in the literature for single image presentations as short as 250ms for monkeys and 350ms for humans suggest that, once taken into account the time needed to produce the motor response, the image is processed in less than 200ms (Fabre-Thorpe et al., 1998; Intraub, 1980; Keysers et al., 2001; Potter, 1976; Rousselet, Fabre-Thorpe, et al., 2002; Rubin and Turano, 1992; S. Thorpe et al., 1996). This is consistent with the fact that both humans and monkeys explore the environment making 2-4 saccades per second.

The quickness and ease of our visual recognition abilities make it hard to appreciate the computational challenge underlying object vision. This became apparent in the early days of computer vision, when Marvin Minsky noted that "*An exploration of the child's world of blocks proved insurmountable, except under the most rigidly constrained circumstances*" (Mead, 1989); the problem itself is rather broad, as objects are recognized and classified not only from their appearance but also by their intended purpose and/or physical properties.

Throughout this thesis we will focus on a specific subset of our visual ability: the assignment to objects of precise (identification) or coarse (categorization) labels that allow to separate them from all other objects, and

the ability to perform this assignment over a variety of so-called identity-preserving transformations (e.g. rotations, translations, scaling, shadowing, background changes, etc.). This set of abilities is known as Core Object Recognition (DiCarlo and Cox, 2007; DiCarlo, Zoccolan, et al., 2012).

The ability to support Core Object Recognition is gradually built in the visual system of primates by increasing the linear separability of the neuronal representations of visual objects. This amounts to assume that a downstream neuron can discriminate the representations of two objects provided by a neuronal population in a given visual area by performing a weighted sum of the synaptic inputs it receives from that area, thus dividing the representations through a hyperplane.

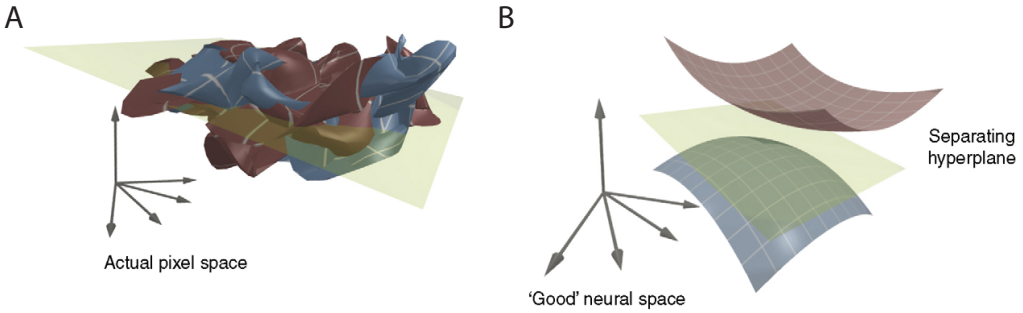

Figure 1.1. Visual object representation of different complexities. Panel **A** portrays a 3D projection of the representation in pixel space of a pair of objects. Panel **B** depicts a representation of the same two objects in a higher level area. Adapted from DiCarlo and Cox, 2007.

In the early visual stages the structures in which the representations of the objects lie (called *object manifolds* (DiCarlo and Cox, 2007)) will be highly curved and tangled with each other, thus making it very difficult to find a hyperplane that is able to separate them (Figure 1.1 **A**). Likely, the goal of the visual system is to gradually reformat these representations, making them flatter and reducing their dimensionality, so as to make it easier to find a hyperplane that separate them (Figure 1.1 **B**).

Such untangling is thought to be performed through a serial cascade of processing modules operating in a largely feedforward manner; this hypothesis stems from the fact that reaction times are very short and incompatible with the involvement of a significant feedback, and from the increasing latencies of the different areas along the so-called ventral visual pathway (see figure 1.4).

1.2 The Primate Visual System

The idea of a hierarchical organization of the visual cortex dates back to the very first days of the study of the functional properties of the visual system. Hubel and Wiesel themselves, when characterizing the properties of receptive fields in cat primary visual cortex, suggested that complex and hypercomplex cells should be built by pooling together the responses of simple and complex cells respectively (Hubel and Wiesel, 1962; 1965).

Anatomical studies indicate that in the macaque monkey there are at least 32 visual areas, accounting for 55% of the neocortex, connected by almost 200 connections, most of which are reciprocal (Felleman and Van Essen, 1991). Figure 1.2 details a scheme of the macaque visual hierarchy, together with the connectivity pattern among the different areas. These many areas are organized into two parallel and distinct processing streams (Figure 1.3), both originating from the primary visual cortex: the dorsal (or parietal) stream, and the ventral (or temporal) stream (Mishkin et al., 1983).

These two streams superintend to the two main processes needed in vision: the dorsal stream, going from V1 to MST along the dorsal-anterior direction, processes visual input to extrapolate the position and direction of motion of the objects in the visual field (leading to the name "the where stream"). The ventral stream, going from V1 to IT along the ventral-anterior direction, processes visual input to determine the shape and identity of whatever items fall in the visual field (and hence the name "the what stream"). Although the two streams are not fully segregated and although the ventral stream hosts other functions in conjunction with object recognition, given its paramount role in supporting invariant object recognition (lesions in IT cortex give rise to very specific deficits in recognition of complex objects, e.g. prosopagnosia (Milner et al., 1991; Munk, 1881; Nicholls et al., 2012; Wada and Yamamoto, 2001)) we will focus our attention on the ventral stream for the rest of this chapter.

As mentioned above, the ventral stream is organized as a hierarchy of areas (1.4). Anatomical evidence of such hierarchical processing stems from the fact that each cortical area is organized in six layers with very well-defined connectivity patterns: layer four receives the feedforward input from the upstream, lower-level area; layers II-III send feedforward projections to downstream, higher-level areas; and, finally, layers 5 and 6 send feedback projections to upstream lower-level areas. By tracking these patterns of connectivity through anterograde and retrograde tracers Felleman and Van Essen (1991) mapped the whole hierarchy of monkey visual cortical areas, producing the diagram shown in Figure 1.2. Another piece of evidence in

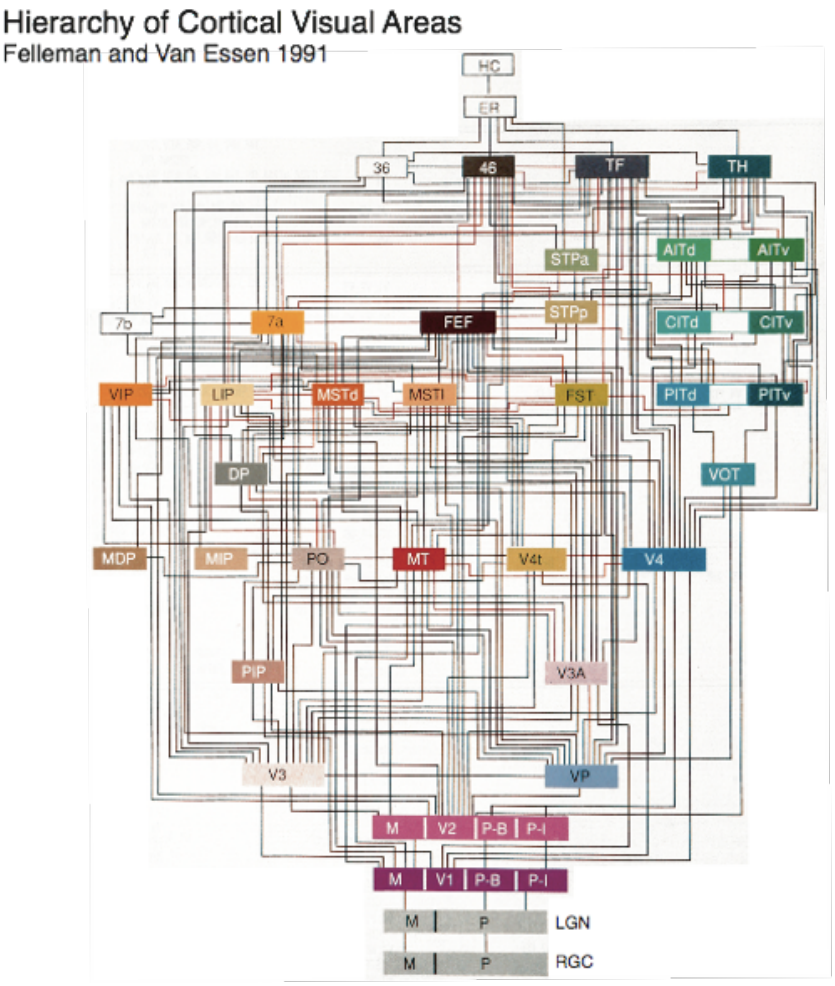

Figure 1.2. Visual hierarchy of the macaque visual system. The different areas are color coded according to their anatomical position. Taken from Felleman and Van Essen (1991)

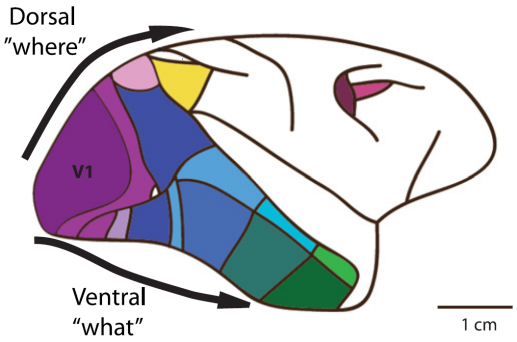

Figure 1.3. The two processing pathways for visual information in the macaque brain. The dorsal pathway extends from V1 towards the upper part of the brain, while the ventral pathway extends towards the lower part. Adapted from Niell (2011)

favor of the hierarchical organization of the visual cortex is the increase in response latency and receptive field size along (e.g.) the progression V1-V2-V4-IT (Figure 1.5)

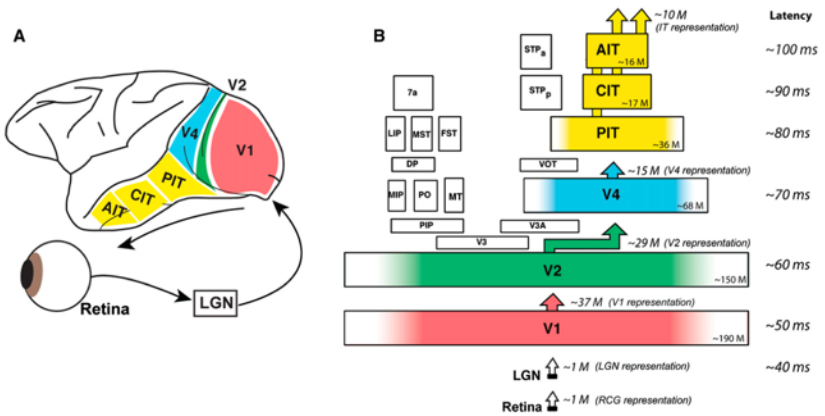

Figure 1.4. Structure of the ventral stream. Panel **A** shows the different areas composing the ventral stream and their position in the brain of the macaque. Panel **B** shows the hierarchical structure of the ventral stream. The size of each area is proportional to the surface area it occupies, and each panel reports the number of neurons in the area (lower right) and the number of outgoing projections (upper right). Adapted from DiCarlo, Zoccolan, et al. (2012)

This increase in latency and RF size is paralleled by an increase in the complexity of stimulus features the neuron is selective for, as higher-level areas of build more and more complex representations of visual stimuli by combining the different features coded by earlier stages. In fact neurons in

V1 have a preference for simple stimuli as oriented edges (Hubel and Wiesel, 1968), while neurons in areas V4 and IT show a marked selectivity for more complex stimuli, such as hands or faces, as shown in figure 1.6 (Desimone et al., 1984; Kobatake and Tanaka, 1994).

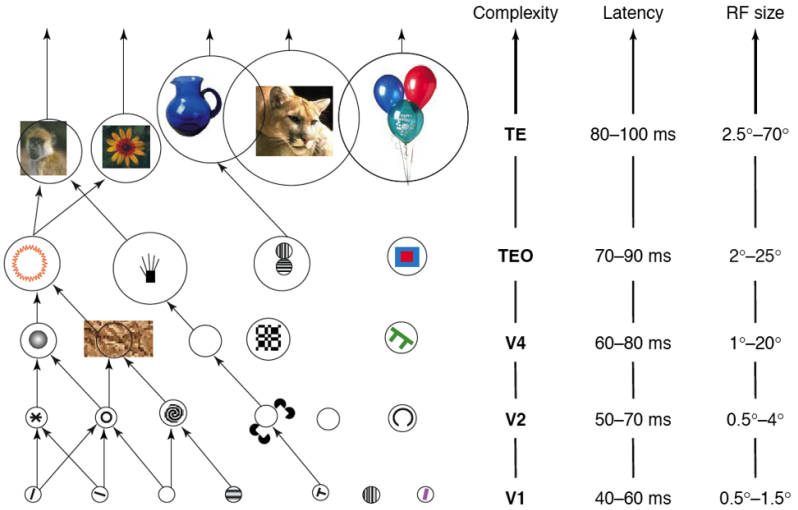

Figure 1.5. Single neuron properties along the ventral stream. The figure shows how moving along the progression V1-V2-V4-IT, due to the successive pooling of lower level outputs, the complexity of the features preferred by the neurons increases, along with response latency and RF size. Adapted from Rousselet, S. J. Thorpe, et al. (2004)

This increase in stimulus selectivity is also paralleled by an increase of tolerance to stimulus transformations, thanks again to the pooling operation performed by downstream areas (Rust and DiCarlo, 2010). These two apparently antithetic requirements, the need for high stimulus selectivity while maintaining a broad tolerance for stimulus transformation, are thought to increase in parallel in a balanced manner in order to disentangle object representations (thus making possible object recognition via simple linear readout), while maintaining a constant sparseness to increase fault tolerance (Rust and DiCarlo, 2012).

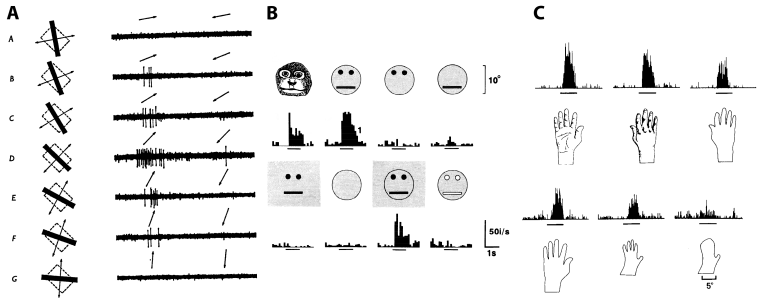

Figure 1.6. Feature selectivity for different areas of monkey visual cortex. Panel **A** shows an example of preferred features for a V1 complex cell (Adapted from Hubel and Wiesel, 1968), while panels **B** and **C** show a face selective cell and a hand selective cell respectively. (Adapted from Kobatake and Tanaka (1994) and Desimone et al. (1984))

1.3 The Rat as a Model of the Mammalian Visual System

Until recently visual neuroscientists focused almost exclusively on macaque monkeys as models to understand object vision and other high-level visual functions. This is due to the fact that nonhuman primates are evolutionarily close to us and have a visual system very similar to our own. Rats and mice, instead, are crepuscular animals and thus heavily rely also on other sensing modalities to perceive the world, seemingly making them unappealing as models for vision research. Moreover their anatomy is quite different from the one of primates: their eyes face laterally and compensate for head movements to maintain a binocular field of view above their head (Wallace et al., 2013); they are afoveate, and their visual acuity is about ten times worse than the acuity of humans (Prusky et al., 2002). (see Figure 1.7)

Nonetheless several behavioral and electrophysiological studies were conducted in the first half of the last century to assess the capabilities of the rat visual system, but, given several limitations of these early approaches (e.g., poor characterization of rat's primary visual cortex, and lack of systematicity), they did not provide convincing evidence that rats are capable of advanced pattern vision (for a review, see Zoccolan, 2015), leaving rat visual cortex an uncharted territory except for a few studies on early visual deprivation (e.g. Tees, 1968; 1972; Tees et al., 1982).

In recent times, however, the dramatic increase in the number of available experimental methodologies and the growing concerns about the use of primates in research have sparked a new interest in the visual abilities of

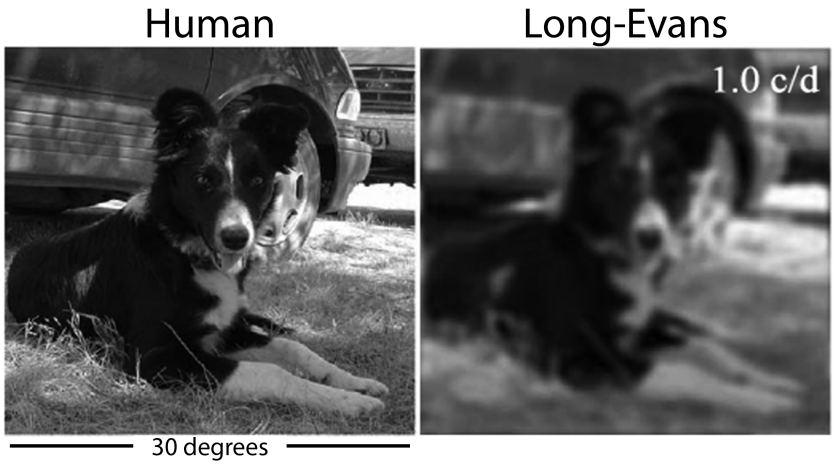

Figure 1.7. Example of an image as it would be seen by the eye of a Long-Evans Norway rat (*Rattus Norvegicus*), the rat strain we used in our study. Adapted from Prusky et al. (2002)

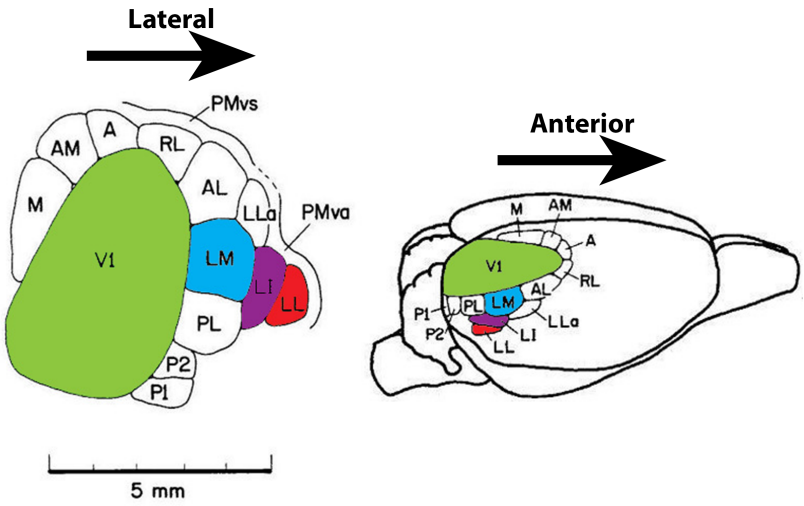

Figure 1.8. Visual Cortical areas in the rat. Highlighted in color are the areas composing the putative rodent ventral stream. Adapted from Sereno and Allman (1991)

rodents. This led several groups to investigate the visual cortex of mice and, to a lesser extent, rats through imaging and electrophysiology experiments (Carandini and Churchland, 2013; Gavornik and Bear, 2014; Glickfeld, Reid, et al., 2014; Huberman and Niell, 2011; Katzner and Weigelt, 2013; Niell, 2015), while other investigators explored rat object vision through behavioral studies. (Reinagel, 2015; Zoccolan, 2015). It was recently found that mice

have at least 13 different visual cortical areas (figure 1.8), the biggest being V1 (occupying about 15% of cortica area) adjoined by many extrastriate areas both in the lateral and medial direction (Sereno and Allman, 1991).

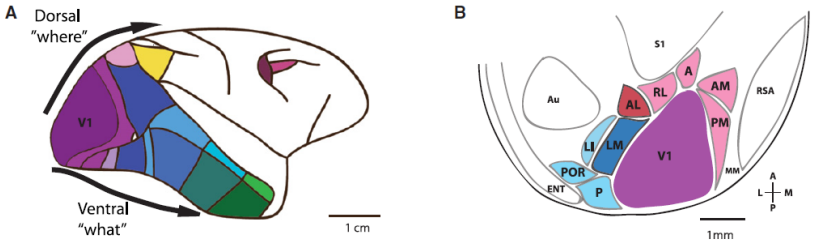

Figure 1.9. Dorsal and ventral stream in mouse visual cortex. The areas composing the dorsal and ventral stream in the macaque visual cortex (panel **A**) are shown together with their homologous in mouse cortex (panel **B**). Corresponding areas are shaded with the same color. Taken from Niell (2011)

Anatomical and physiological studies in mice have revealed that these areas are organized in a hierarchy (Glickfeld and Olsen, 2017) similar to the one of the macaque cortex (figure 1.9). In particular functional imaging studies suggest that mice medial and lateral visual areas are analogous to the areas belonging to the monkey dorsal and ventral streams, respectively (Niell, 2011).

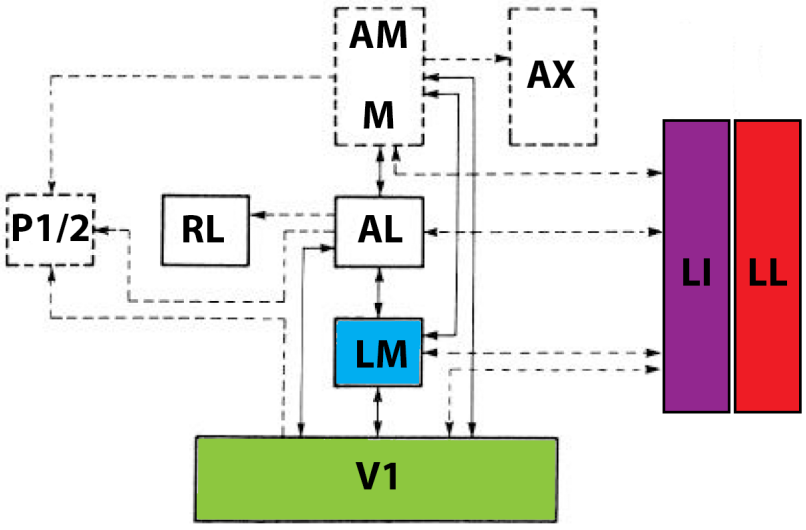

Figure 1.10. Hierarchical organization of rat visual cortex. Areas belonging to the putative ventral stream are shaded with the same color code as in figure 1.8. Adapted from Coogan and Burkhalter (1993).

To what extent rodent visual cortex was capable of supporting high-level processing of visual object information was mainly investigated in the rat, which, similarly to the mouse, has a visual cortex organized in a hierarchical fashion (Coogan and Burkhalter, 1993), with each area having a well-defined retinotopic organization (Espinoza and Thomas, 1983). Recent behavioral and electrophysiological evidence indicates that rats are able to recognize objects in spite of identity preserving transformations and that rat lateral extrastriate areas, thought to compose the putative rat homologous of the primate ventral stream (see fig. 1.8 and 1.10), gradually prune visual representations from low-level information in order to code for more complex features in an invariant way (Alemi-Neissi et al., 2013; Tafazoli, Di Filippo, et al., 2012; 2014; Tafazoli, Safaai, et al., 2017; Zoccolan, 2015; Zoccolan et al., 2009).

In my PhD work I have further probed the object recognition capabilities of rat visual cortex by studying how the information flows across the putative ventral stream of the rat, and by looking at whether the representations of visual objects change to support invariant object recognition. As explained in the Results, the results of this investigation will add to the aforementioned body of evidence supporting the rat as a model of higher visual processing.

Chapter 2

Materials and Methods

As explained before, the rat is gaining prominence as a model of the mammalian visual system. To further corroborate the behavioral findings, and to understand the mechanisms underlying rat's object recognition capabilities, we recorded single neuron and population responses to the presentation of visual stimuli across all the areas that are thought to compose the rat ventral stream.

We will now detail the methodologies we used and the analyses we carried out to probe the neural representation of visual objects.

2.1 Experimental Design

2.1.1 The Stimulus Set

For our experiments we built a rich and ecological stimulus set using a large number of objects, organized in a semantic hierarchy (figure 2.1). We chose to organize the stimulus set according to a semantic tree not because we expect rat visual cortex to be specialized for the representation of specific classes of natural objects. Rather, we exploit semantic similarity as a proxy for visual/structural similarity. This allows sampling densely objects sharing a common set of features (e.g., four-leg animals), while at the same time spanning various corners of the shape space, given the variety of shapes/textures of our semantic classes.

To build the stimulus set we used a set of 40 3D models of real world objects (TurboSquid, New Orleans, LA, USA), both natural and artificial, each rendered in 36 different poses, randomly chosen around four main views (frontal, lateral, top, and 45° in azimuth and elevation) for a total of 1440 stimuli. In addition these poses underwent also a set a position and size

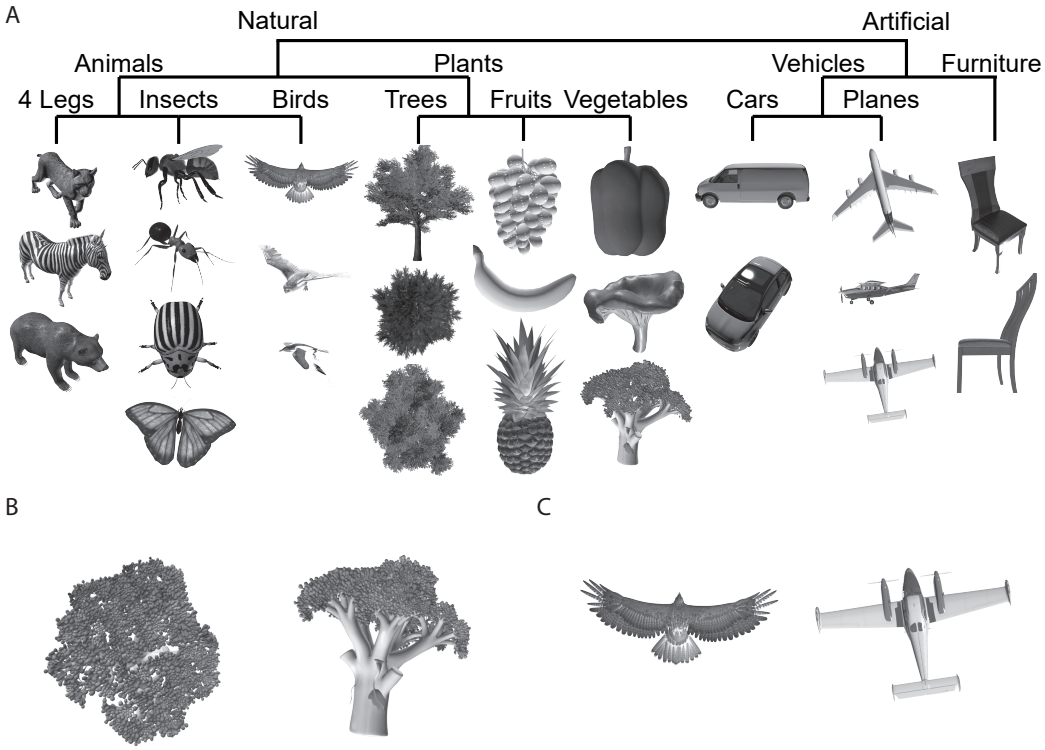

Figure 2.1. Tree structure of the objects composing the stimulus set (panel **A**). It can be clearly seen how the objects present similarities and differences both within and between categories. As an example, the different poses of a broccoli (panel **B**) can lead to drastically different images, while birds and planes (panel **C**) share some distinctive features (e.g. the wings)

changes (see next section for details). The stimuli were rendered using 3D Studio Max (Autodesk, Inc., San Rafael, CA, USA), a commercially available 3D rendering software.

Each stimulus was presented over grey background at one of three possible sizes (30°, 35°, and 40°) chosen at random, in one of three possible positions (0°, ±15°), also chosen at random, and rotated in plane of either 0°, 90°, or ±45°. Note that these transformations add to the pose variations produced when rendering the 3D models. To increase the amount of variability of the appearance of the objects, a small amount of random noise, of about 5° of amplitude, was also added to all the above mentioned transformations (e.g., 5 translation, scaling or in-plane rotation).

After rendering we analyzed the screenshots of the stimuli, exactly as they were presented during the protocol. Through a combination of custom and built-in software written in MATLAB (The Mathworks, Inc., Natick, MA, USA) we characterized the visual statistics of the stimulus set by extracting the bounding ellipse, the perimeter and the values of the pixel intensity (we used for this purpose the built-in function regionprops) and from these we derived several parameters.

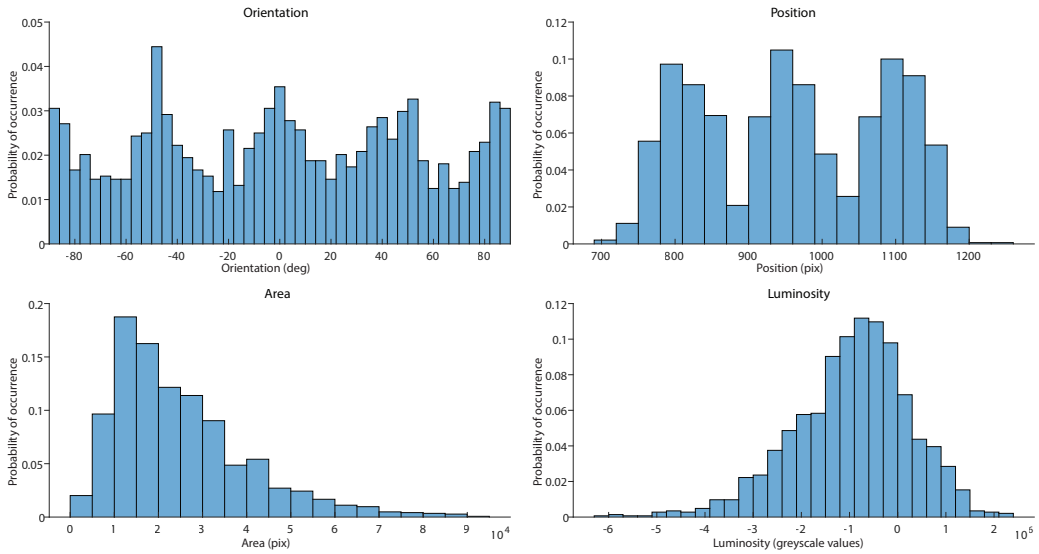

Figure 2.2. Histograms of the distribution of the parameters of the stimulus set. In the panel we show the distributions for the orientation, position, area, and luminosity.

Specifically we extracted a set of low and mid level features (figure 2.2), for the images of the stimuli:

- Position for the center of mass

- Luminosity (defined in various ways; see below)
- Area
- Contrast
- Aspect Ratio
- Orientation

From the histograms in figure 2.2 it can be clearly seen how there are three peaks for the position (corresponding to the three main presentation positions) and four peaks for the orientation (corresponding to the four angles of rotation applied, considering ±90 as the same oreientation).

We defined total luminosity L_{Total}, RF luminosity L_{RF}, and contrast Σ_{RMS}, respectively, as

$$L_{Total} = \sum_{x=1}^{1920} \sum_{y=1}^{1080} I(x, y) \tag{2.1}$$

$$L_{RF} = \sum_{x=1}^{1920} \sum_{y=1}^{1080} I(x, y) \cdot RF(x, y) \tag{2.2}$$

$$\Sigma_{RMS} = \sqrt{\frac{1}{1920 \cdot 1080} \sum_{x=1}^{1920} \sum_{y=1}^{1080} (I(x, y) - \bar{I})^2} \tag{2.3}$$

where $I(x, y)$ is the intensity value of the pixel in position (x, y), $RF(x, y)$ is the normalized receptive field of the neuron in position (x, y), while $\bar{I}$ is the mean intensity across all pixels.

2.1.2 The Presentation Paradigm

Before the presentation of the protocol we ran a separate block to map the receptive fields. In this protocol moving bars 10° wide were shown on a black background at different positions and at 8 different orientations (0°, 45°, 90°, 135°, 180°, 225°, 270°, 315°), dividing the screen in a 11 × 6 grid (Niell and Stryker, 2008; Tafazoli, Safaai, et al., 2017). From this first block we estimated online the positions of the receptive fields (RFs) and by observing their progression along the azimuth (in particular, the points of inversion) we identified the areas reached by the electrode array, and we assigned each recording channel to a specific visual cortical area (Espinoza and Thomas, 1983). After this first block of recording ended, we proceeded to administer

the main protocol, interleaving it with the RF mapping procedure to measure the receptive fields of the neurons that were tested with areaour battery of object conditions.

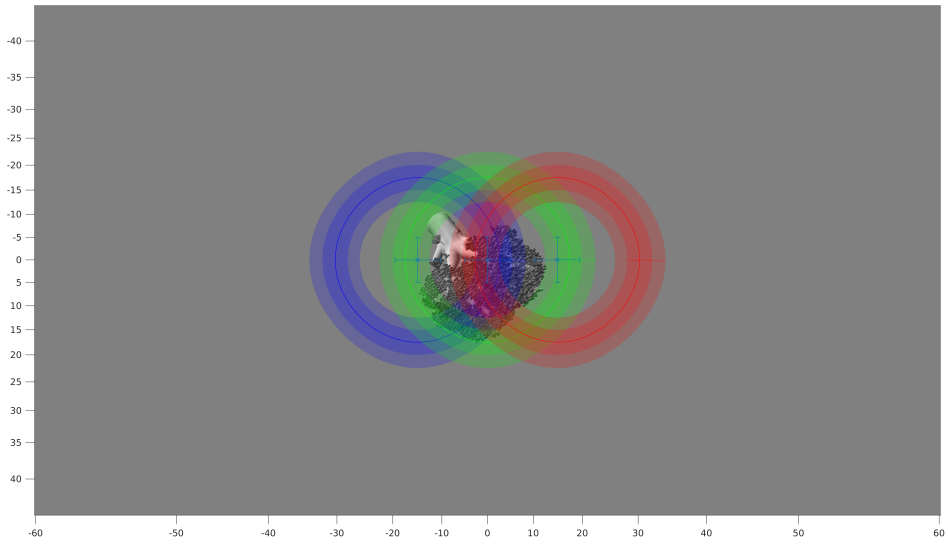

Figure 2.3. Schema of the presentation paradigm. The three coronas represent the three standard sizes the stimuli have, while the shaded area represent the random 5° noise added to it.

Each stimulus was presented for 150ms and was followed by a 250ms interstimulus interval, in a protocol known as Rapid Serial Visual Presentation (RSVP). This method allowed us to minimize the impact of feedback connections, focusing on the initial swept of activity that propagates through the processing hierarchy in a feed-forward way. This also allowed keeping the protocol running time within 2 hours (which important to increase the likelihood of having a stable recording of single neurons), all while having a large number of presented stimuli and without compromising the tuning and selectivity of neurons in higher-level areas (Földiák et al., 2004; Keysers et al., 2001).

2.2 Surgical Procedures

All procedures involving animals were conducted in agreement with international and institutional standards for the care and use of animals in research and were approved by the Italian Ministry of Health.

We used Naïve Long-Evans male rats (from Charles River Laboratories International, Inc.) to record extracellular action potentials from the primary visual cortex (V1) and from the three lateral extrastriate areas that compose the first stages of the rat's putative ventral stream, namely Lateromedial (LM), Lateroinermediate (LI), and Laterolateral (LL). All animals had fully completed the development of their visual system at the time of the surgery, (Berardi et al., 2000) and weighted between 350 and 500 grams.

Anesthesia was induced with a solution of 0.3 mg/kg fentanyl (Fentanest, Pfizer, 0,1 mg/2 ml) and 0.3 mg/kg medetomidin (Domitor, Orion Pharma, 1 mg/ml) solution, administered via intraperitoneal injection. Rats were kept in a state of deep sedation, and the level of anesthesia was monitored throughout the procedure by checking the absence of tail and paw reflexes; after an hour from the induction of anesthesia, we switched to the maintenance regime (0.1 mg/kg fentanyl and 0.1 mg/kg medetomidin) to maintain the anesthesia level. Body temperature was kept stable at 37°C using a heated pad (Harvard Apparatus, Holliston, MA, USA) and a steady supply of oxygen was delivered to the animal's nose to counterbalance the opioid-induced respiratory depression. During the whole operation the rat's eyes were kept moisturized using an ophthalmic solution (EPIGEL from Ceva Santé Animale, Libourne, FR)

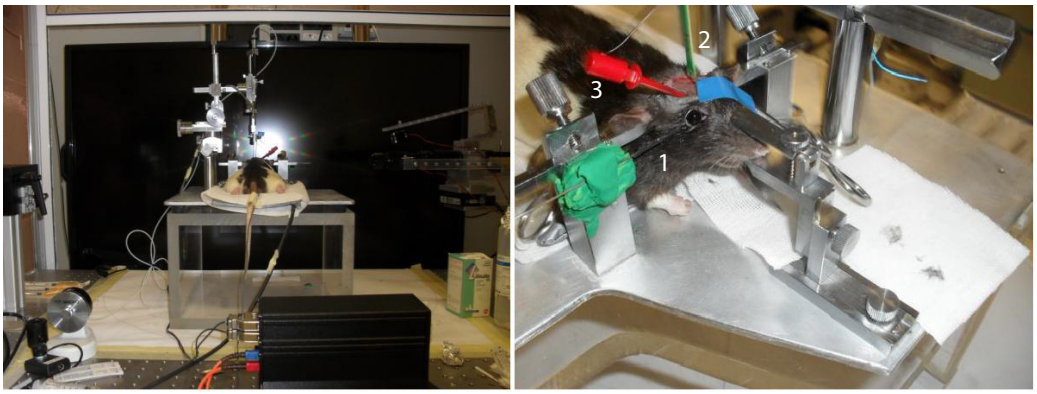

Figure 2.4. View of the recording setup. A) Stereotaxic apparatus and elevated platform on which it is placed during recording. The platform is 21 cm high and is placed on a rotating stage in order to center receptive fields on the screen. B) Detail of the head of the animal. The eye-ring (1) keeping the eyelid open is fixed to the apparatus, while the electrode (2) is mounted on the arm, ready to be inserted into the brain. The clip at the edge of the skin incision (3) serves as ground reference. Adapted from Tafazoli (2013)

After anesthesia induction the animal was placed on a stereotaxic apparatus to take the appropriate measurements for the craniotomy coordinates.

All the recordings were carried out on the left hemisphere to account for the necessary presence of the arm of the stereotaxic apparatus, while the left eye was kept closed in order to avoid possible interference coming from the light entering the contralateral eye; coordinates for the craniotomies were extrapolated from the areas found in Paxinos and Watson (2007) and from the results of previous surgeries conducted in the lab.

Once opened the craniotomy at the intended site, the dura mater was mechanically removed and the brain was kept hydrated using physiological saline solution (0.9% NaCl in H_2O) for the whole duration of the surgery. The stereotaxic apparatus was then placed on an elevated platform (Fig. 2.4) in order to keep the eye pointed at the center of the screen at a distance of 30cm, and the right eye was kept fixed using an eye-ring; at this point the electrode array was coated with a DiI solution and inserted into the brain, so as to allow the histological reconstruction post-mortem of the insertion track.

2.3 Electrophysiology

Neural activity was recorded using linear 64 channel silicon electrodes arrays (NeuroNexus Technologies, Ann Arbor, MI, USA) with different designs according to the area targeted in each surgery. Given the better accessibility of primary visual cortex (V1) we used 4 shanks probes (figure 2.5, right) with either 5mm or 6mm length, $50\mu m$ site spacing, $177\mu m^2$ site area and $200\mu m$ shank spacing (models $\mathrm{A4} \times 8 - 5\mathrm{mm} - 50 - 200 - 177 - \mathrm{A32}$ and $\mathrm{A4} \times 16 - 6\mathrm{mm} - 50 - 200 - 177 - \mathrm{A64}$) to target it, as it allowed us to better span the visual field. Instead, to record from lateral extrastriate areas we used either single shank or two shanks electrodes (figure 2.5, left) with either $50\mu m$ or $25\mu m$ site spacing, $177\mu m^2$ site area and $200\mu m$ shank spacing for the two shank electrodes (models $\mathrm{A1} \times 32 - 6\mathrm{mm} - 50 - 177 - \mathrm{A32}$ and $\mathrm{A2} \times 32 - 5\mathrm{mm} - 25 - 200 - 177 - \mathrm{A64}$), as they allowed us to record from multiple areas simultaneously and to reach the deeper and more lateral areas (LI and LL).

Activity was recorded using an RZ2 BioAmp signal processor (Tucker-Davis Technologies, Alachua, FL, USA), with a sampling rate of 24.414 kHz.

After inserting the electrode at the desired depth, we carefully mapped the RF positions on the screen using moving bars presented for 200ms, with an interstimulus interval of 200ms, on a widescreen full-HD monitor (SHARP PN-E471R, 1920X1080 pixel resolution; 60 Hz refresh rate; 9ms response time; 700 cd/m^2 maximal brightness; 1200:1 contrast ratio). Following this RF mapping procedure, we rotated the platform holding the stereotaxic ap-

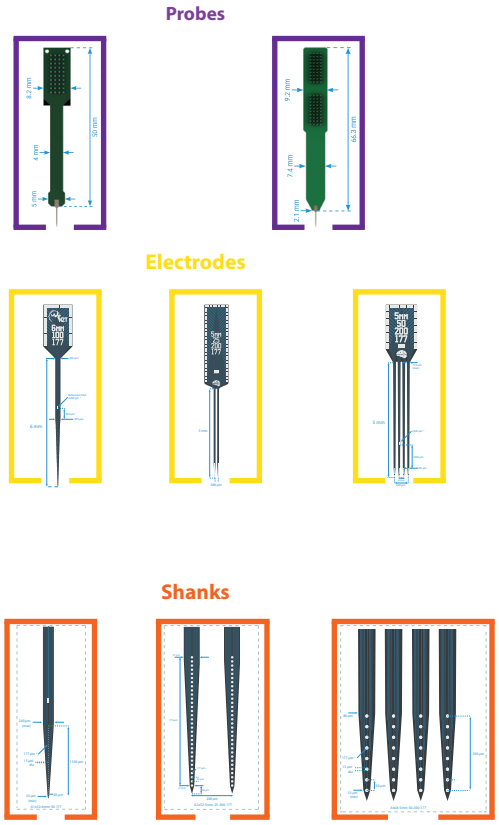

Figure 2.5. Diagram detailing the configurations of the electrodes used in the recordings.

paratus, so as to put the receptive fields at the center of the screen. Finally, we proceeded to estimate the area(s) reached by the array by looking at the well-established reversal of the retinotopy that takes place at the borders of adjacent visual cortical areas (Espinoza and Thomas, 1983).

2.4 Data Analysis

2.4.1 Spike Sorting

Once collected, raw extracellular waveforms were analyzed using the KlustaKwik spike sorting suite (Rossant et al., 2016). The waveforms were filtered to remove the local field potential and then were processed through a masked Expectation-Maximization algorithm that took into account probe geometry in order to better separate spikes belonging to different neurons.

The results of the automatic spike sorting procedure were then manually refined using a specialized tool (phy,) to merge clusters that were incorrectly labeled as different or to merge clusters of spikes that were separated due to evident external causes (e.g. electrode drift or momentary bursting activity). At the end of the spike sorting procedure, all clusters were labeled either as good single units or as multiunits according to the following criteria: 1) spike trains autocorrelograms; and 2) refractory period violations, labeling as multiunits those clusters that had more than 0.5% of interspike intervals smaller than 1ms.

2.4.2 Data Preprocessing

As a first necessary step towards subsequent analyses, we preprocessed the raw spike times coming from the spike sorting procedure to extract spike count windows and spike counts.

To estimate the latency of each neuron we followed the procedure detailed in Baldassi et al. (2013). We first computed the firing rate of the neuron for each stimulus presentation by convolving the pointwise distribution of spike times with a square window 25ms wide. This yielded peri-stimulus time histograms (PSTHs) for each stimulus conditions. We then averaged the PSTHs of the 10 best (i.e., most effective) stimuli, which were selected by computing the overall spike count over a fixed 200ms wide tentative window starting at 25ms from stimulus onset (figure 2.6, panel **B**). Finally, we subtracted the baseline(or background) activity from the resulting average PSTH, with the baseline being defined as the average firing rate in the 50ms preceding the stimulus onset. From the resulting driven rate profile, we defined the starting

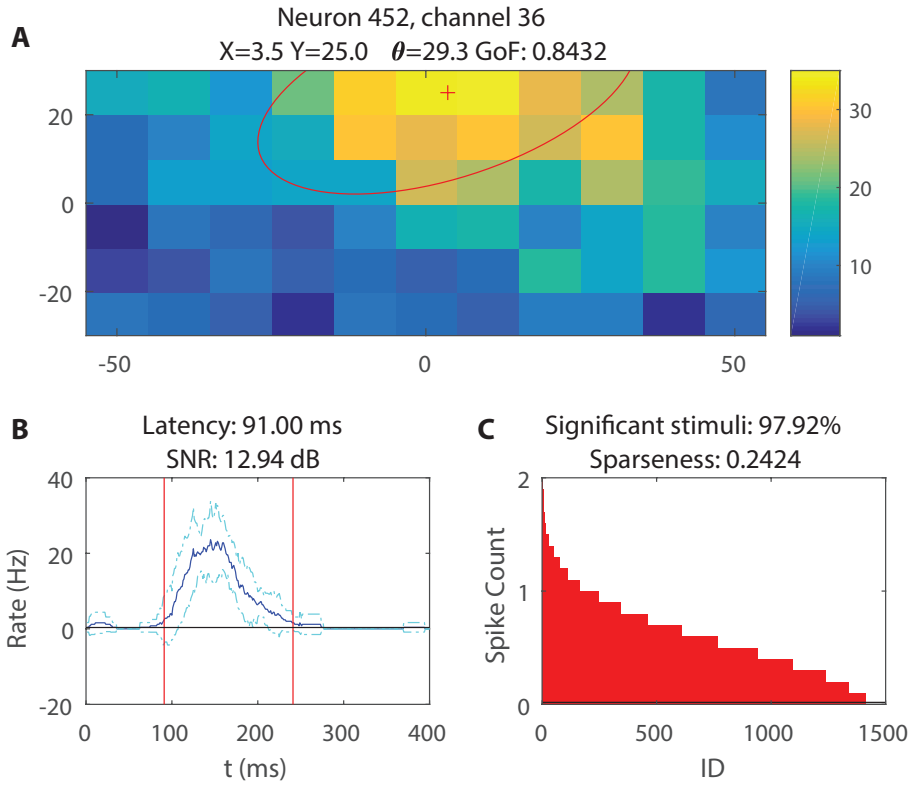

Figure 2.6. Summary of data preprocessing for an example neuron. in panel A the activity from receptive field mapping trials is shown, together with its best-fitting Gaussian profile (1σ). Panel B shows the firing rate profile averaged over the 10 best trials. Light blue lines mark the beginning and the end of the spike count window. In panel C we report the rank-ordered plot for the example neuron and its computed sparseness. Red bars mark stimuli with significant activity (significance was assessed with a two-tailed permutation test).

point of the spike count window (i.e. the latency of that neuron) as the time at which the firing rate first crossed a threshold fixed at 20% of the peak firing rate; the window had a fixed width of 150ms.

We then computed receptive field position and size of the recorded neurons by fitting the raw RF maps (see figure 2.6, panel **A**) to a 2-dimensional Gaussian function with unequal variances, according to the following equations:

$$\begin{aligned} r &= a \cdot e^{-\left(\frac{\hat{x}^2}{2\sigma_{\hat{x}}^2}+\frac{\hat{y}^2}{\sigma_{\hat{y}}^2}\right)} + b \\ \hat{x} &= (x - x_0)\cos(\theta) - (y - y_0)\sin(\theta) \\ \hat{y} &= (x - x_0)\sin(\theta) + (y - y_0)\cos(\theta) \end{aligned} \tag{2.4}$$

where a is the peak amplitude of the Gaussian, b is its offset, (σ_x, σ_y) is its variance along the two axes, θ is its angle with respect to the horizontal, and (x_0, y_0) is its center.

The resulting receptive fields were then used do compute the receptive field luminance (as detailed in Tafazoli, Safaai, et al., 2017) and to assign the neuron to its corresponding area by looking at the receptive field progression.

2.4.3 Single Cell Analyses

Information Theory

In order to give understand the inner workings of a system, it is important to understand the relations that link the input to that system to its output. These relation unfortunately can (and will) be highly nonlinear, especially in the case of electrophysiological data where the input to a neuron is the result of a cascade of processing modules that are not directly accessible.

Mutual Information is a non-linear measure that quantifies the amount of information that can be extracted from a variable (namely, the features of a stimulus) given the fact that we know another variable (namely, the neuronal response). Comparing the conditional probabilities $P(r|s)$ of observing a response r when a stimulus s is presented to the unconditional probability $P(r)$ of observing a response r under any condition, we can derive a quantity that tells us how well an ideal observer can discriminate between stimulus conditions given the response to a single trial presentation (Borst and Theunissen, 1999).

Mathematically, mutual information is computed as:

$$I(R;S) = \sum_S P(s) \sum_R P(r|s) \log_2 \frac{P(r|s)}{P(r)} \tag{2.5}$$

where $P(s)$ is the probability of seeing a stimulus s, $P(r|s)$ is the probability of seeing a response r to a stimulus s, and $P(r)$ is the probability of seeing a response r in any condition.

Since some of the features that characterize the objects in our stimulus set can be parameterized, it is interesting to compute the mutual information between the neural response and the different features of the stimuli. This information is computed as

$$I(R; F_i) = \sum_{F_i} P(f_i) \sum_{R} P(r|f_i) \log_2 \frac{P(r|f_i)}{P(r)} \tag{2.6}$$

where f_i is the value of the feature F_i. It would be tempting to assert that if $I(R; F_i) > 0$ the neuron codes for the feature F_i, but unfortunately things are not so simple when features are not independent, i.e. when $I(F_i, F_j) > 0$. In this case, a positive value of $I(R; F_i)$ could simply be due to the fact that F_i and F_j spontaneously co-occur, and not because of a specific encoding of F_i by the neuron.

We addressed this problem by separately computing the mutual information between neural response and a feature when all the others are kept at constant value. This quantity is called Conditional Mutual Information (Ince et al., 2012) and is defined as follows:

$$I(R; F_j|F_i) = I(R; F_j \ \& \ F_i) - I(R; F_i) \tag{2.7}$$

It is natural, in our case, to use, as features, those that we already measured for our stimulus set; for example F_i could be the position of the stimulus (left, right, or center), while F_j could be its orientation with respect to the horizontal axis. In this case $I(R; F_j|F_i)$ would represent the amount of information about the orientation of the stimulus conveyed by the neuronal response which is independent from its position (i.e. at fixed position).

With regard to the details about how mutual information was computed, in order to perform the computation of the sum for continuous quantities without turning it into an integral, we discretized the continuous variables in meaningful bins. We chose the binning keeping into account the distributions of the parameters, so that the division was reasonable; as an example we chose to divide the x position in three bins centered on the three main presentation positions, that correspond to the three peaks in the parameter histogram (see figure 2.2). Moreover, all the binnings were chosen in such a way to keep at least 5 trials for every condition, as it has been proved that by applying the Panzeri-Treves correction (Panzeri, Senatore, et al., 2007; Panzeri and Treves, 1996) even such a small number of trials is sufficient to avoid limited sample bias.

All computations were performed using a combination of custom software and the Information Breakdown ToolBox (IbTB) and were corrected for limited sample bias using the Panzeri-Treves correction (Magri et al., 2009; Panzeri and Treves, 1996).

2.4.4 Population Analyses

The Population Vector Space

Having explored the dynamics of single neuron responses, we moved to investigate the behavior of the different areas of rat's visual cortex at the level of whole populations. To this extent it is necessary to introduce the concept of population vector space.

If we take a population of N neurons, we can fully describe its state at any time point by specifying each neuron's instantaneous firing rate $\nu_1, \nu_2, \cdots, \nu_N$. Considering the population as a whole, the state of our population will become a vector $\nu = (\nu_1, \nu_2, \cdots, \nu_N)$ living in a vector space $\mathbb{V}$ that is a subset of $\mathbb{R}^N$ (the firing rate is positive definite and has a theoretical superior limit of about 1kHz).

In this framework, the evolution of the activity of a population in time will simply be a trajectory in $\mathbb{V}$ while the population response to a stimulus will correspondingly be represented by a point. If we now take all the population responses belonging to a single object undergoing some arbitrary continuous transformation (e.g. rotations), they will lie in a manifold in $\mathbb{V}$ (figure 2.7). The ability of a population to support invariant recognition with respect to the aforementioned transformation can now be defined by the possibility of building an hyperplane that separates the manifold belonging to one stimulus from all the others.

In general, since an object seen under different identity preserving transformations (e.g. rotations, scaling, background changes, etc...) will cast radically different images on the retina, it is quite reasonable to assume that the representations of different visual objects in low-level areas are going to lie in highly curved manifolds entangled between each other (figure 2.7, panel **A**). It follows that, in proceeding upwards across the hierarchy of visual areas, the representations at different levels should become more and more separable, meaning that each area should progressively disentangle the representation manifolds belonging to different objects (figure 2.7, panel **B**).

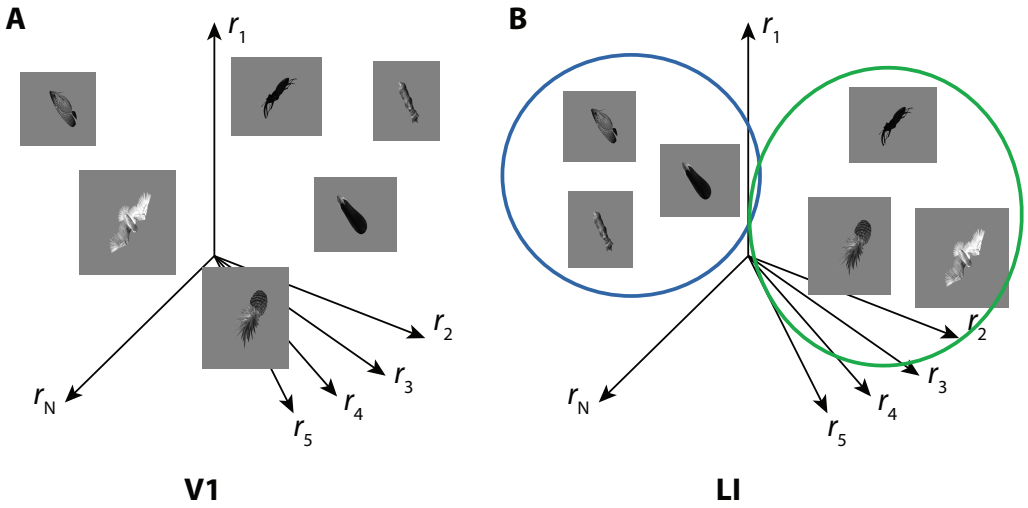

Figure 2.7. Object representation in neural population space. On the left is depicted the population vector space of a low-level visual area. The representations belonging to different objects are tangled and mixed with each other, thus making it very hard, if not impossible altogether, to separate the views of one stimulus from those of another. The situation is different in a higher-level area (on the right panel), where the representations belonging to a stimulus lie close together and well separated from all the others. In this case the object of interest is easily separable from all the other possible stimuli.

Representational Dissimilarity Matrix

When analyzing the responses of a population of neurons to a set of stimuli, it is useful to know which stimuli are considered similar by the population in order to infer the features, shared by the stimuli, that are encoded by the neuronal ensemble. A useful approach in this regard is to build a Representational Dissimilarity Matrix, as done, for instance, by Kriegeskorte, Mur, et al. (2008). Such a matrix is built by computing the distance between the population response vectors of each pair of stimulus conditions. Figure 2.8 shows a scheme detailing how to build one.

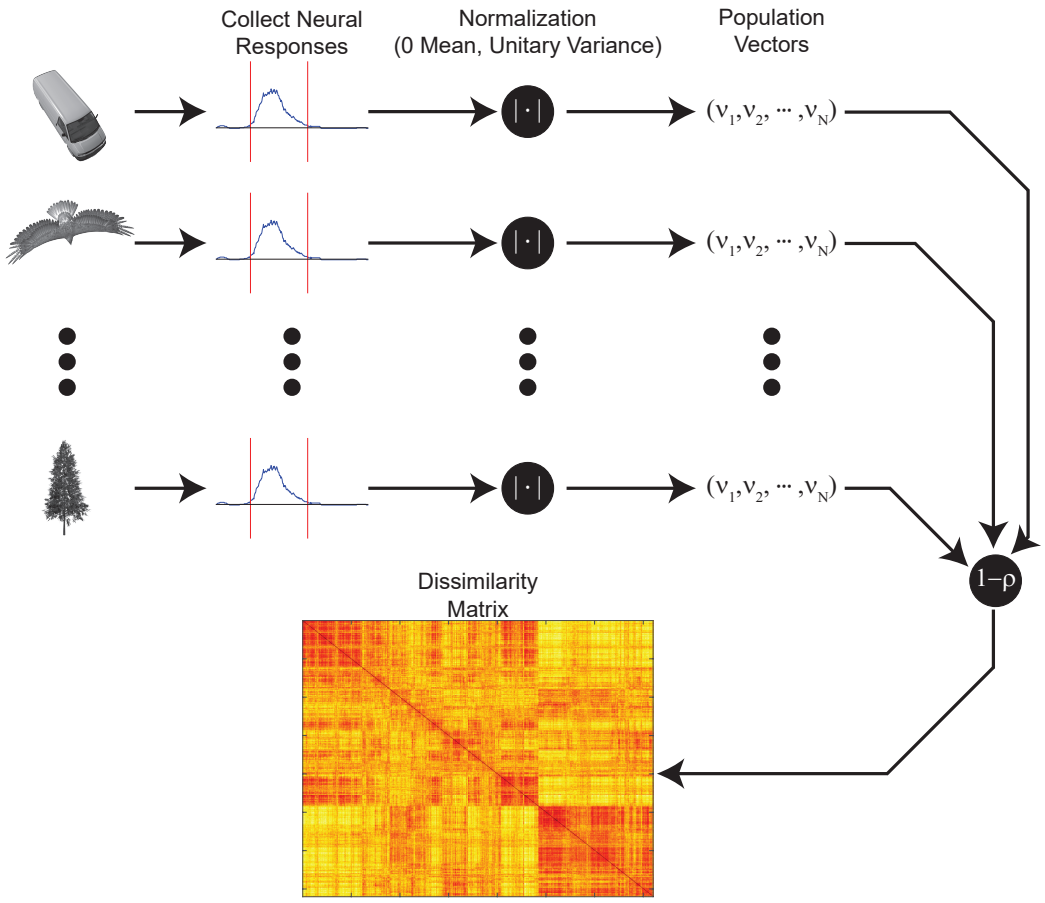

Figure 2.8. This figure details the procedure necessary to build a Representational Dissimilarity Matrix. First, neuronal responses (from single units, MUA, fMRI, EEG, etc...) are collected and normalized. Then a distance (e.g. correlation distance) is computed between each pair of stimuli and arranged in a matrix. Finally a dendrogram of the stimulus set is built and the matrix itself is rearranged according to it, making the underlying structure emerge.

To be able to compute the dissimilarity matrix we first had to choose a metrics to compute distances. Following Kriegeskorte, Mur, et al. (2008) and Kriegeskorte and Kievit (2013), we normalized the population response vectors by subtracting the mean and scaling to unit variance, and then computed the correlation distance defined as:

$$M_{ij} = 1 - \rho(\nu^i, \nu^j) \tag{2.8}$$

Once computed the distance matrix, we used it to build a dendrogram (i.e., a hierarchical tree) of the stimulus set using the complete linkage hierarchical clustering method. In this way we could rearrange the rows and columns of the Representational Dissimilarity Matrix so that similar stimuli (in the neuronal representation) are close together. At the end of this procedure, the structure hidden in the representation of the stimulus set emerges visually in an intuitive way, giving us a tool to guide subsequent analyses.

Quantification of visual features along the hierarchical clusters

Once computed the dendrogram we went on to analyze how the different parameters that define our stimulus set were distributed in the clusters produced by the algorithm, and how they varied as a function of the depth of the tree.

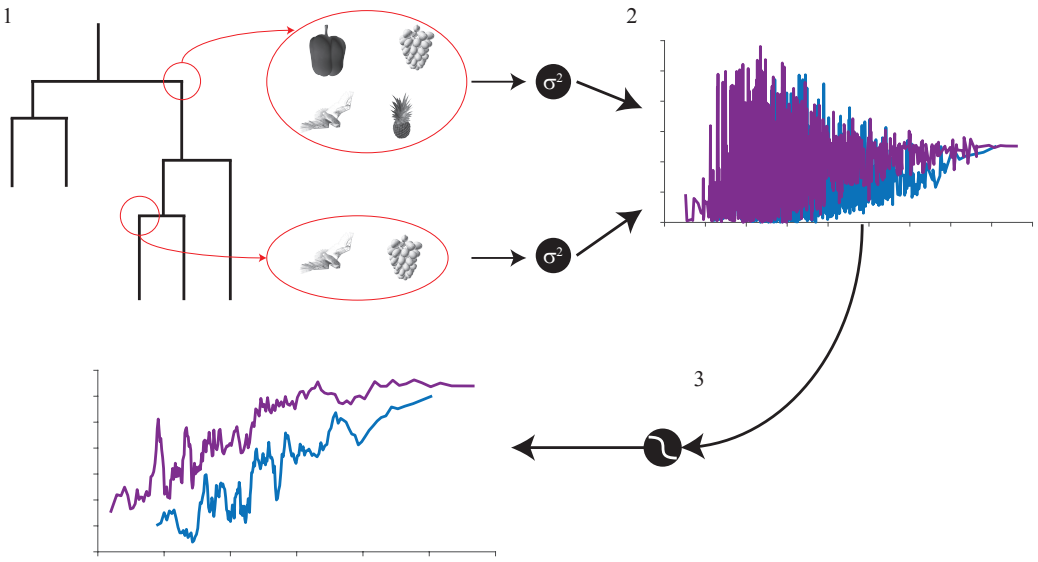

Figure 2.9. Analysis of the evolution of stimulus parameter along stimulus set dendrogram. For each node of the dendrogram we computed the variance of a chosen parameter over the stimuli belonging to that node. The resulting values are plotted as a function of node height and then smoothed.

The rationale behind this analysis is simple: if an area groups the stimuli according to the value of a certain parameter (let's say position along the x axis), stimuli having close values of that parameter will lie near each other in the (small) deepest clusters (or nodes) of the tree (the leaves). On the other hand, the large clusters near to the root of the tree will contain a large number of stimuli, spanning a wide range of values of the property under consideration.

To quantify this, we computed the variance of such property (e.g., position) across all the stimuli falling under a node, repeated over all the nodes and plotted as a function of the depth of the node; since the resulting signal was very noisy, we also smoothed it using a moving average lowpass filter (see figure 2.9). Our expectation is that the variance should exhibit a sharper and faster decrease as a function of the depth of the tree for all the properties a visual area is selective for, while the variance should decrease more gently (if at all) for all the visual features that are not represented in the area.

Dimensionality Analysis

As we have seen earlier, the neuronal responses of a population of N neurons live in an N-dimensional space $\mathbb{V}$; the representations of a stimulus, however, live on a lower-dimensional manifold in this space. Now since we know that the representation manifolds must be progressively smoothed and disentangled along the visual areas' hierarchy in order for it to support invariant object recognition, it is reasonable to deduct that their dimensionality should accordingly decrease.

We estimated the dimensionality of the representation of our dataset in the four different areas using a novel method developed by the group of Prof. Laio (department of Molecular and Statistical Biophysics, SISSA, Trieste, Italy). For each element of our stimulus set we compute the Euclidean distance from its first and second neighbor, and their ratio μ. In the approximation of locally uniform density, the probability distribution function for the values of μ and its cumulative can be expressed as (Facco et al., 2017):

$$\begin{aligned} f(\mu) &= d\mu^{-d-1} 1_{[1,+\infty]}(\mu) \\ F(\mu) &= (1-\mu^{-d}) 1_{[1,+\infty]}(\mu) \end{aligned} \tag{2.9}$$

from which we get, by computing the logarithm of both sides

$$\frac{\log(1-F(\mu))}{\log(\mu)} = -d \tag{2.10}$$

The dimension can thus be estimated by plotting the values of $\log(\mu)$ and $-\log(1-F(\mu))$ and fitting them with a line passing through the origin. This

procedure however has the problem of overestimating the dimensionality of the data, since the noise inherently present in the measurement will randomly distribute the points in the whole vicinity of their low dimensional manifold, thus misleading the algorithm.

To avoid this problem a block analysis has to be performed. First, the dataset is sequentially shrinked in order to increase the probability that the two nearest neighbors of a point will lie in the manifold where the dataset lives. Then the dimensionality of each of these reduced datasets is computed and plotted as a function of dataset size. It is obvious that an excessive reduction of the dataset will lead to a severe underestimate of the dimensionality because the two nearest neighbors will lie too far from the point, but there is a region where the dataset reduction will have mitigated the effects of the noise while keeping mostly intact the low-dimensional structure f the data. True dimensionality will then appear as a plateau near the size of the full dataset (see figure 2.10).

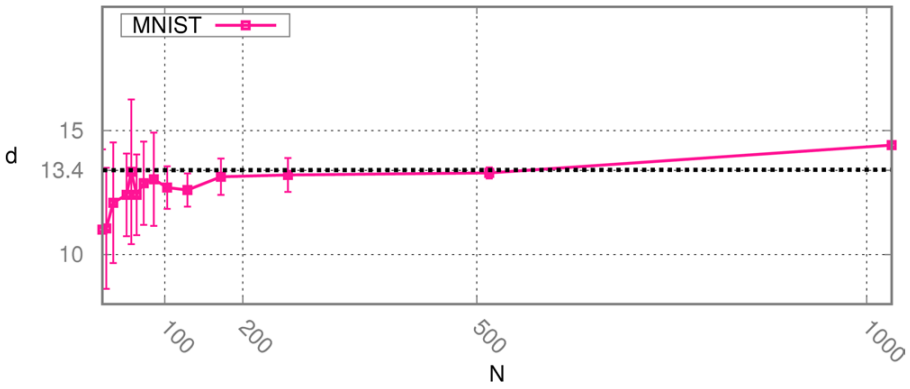

Figure 2.10. Example of the block analysis computed on the digit 2 of the MNIST database. Near the end of the plot a plateau is clearly visible, indicating the true dimensionality of the dataset as 13.4. Picture taken from Facco et al. (2017)

Dimensionality Reduction

In order to assess whether there was a component of the population response driven by specific stimulus features, and to be able to quantify the relative importance of such components we tried to reduce the dimensionality of our dataset by applying Principal Component Analysis.

Principal Component Analysis transforms the data into a new coordinate system chosen in such a way that the first coordinate (also called first principal component) is the one that captures the greatest variance of the data,

the second component is the one that captures the second greatest variance, and so on. This means that the order of the principal components is related to the amount of information that the component carries about the dataset. To investigate to what extent the population response was driven by different parameters in the different areas, we computed the correlation between each parameter and each principal component and then we plotted the values of the two mostly correlated principal components using a color code to report the values of the parameter of choice.

In this way we obtained an easily interpretable map between parameters and principal components, as the presence of a dependency between the parameter and the principal component resulted in a clear and visible color gradient. Moreover, the efficacy of a parameter in driving the population response can be quantified both by the amplitude of the correlation coefficient and by the explained variance of the principal component it correlates the most with.

K-Means Clustering

With the last analysis we performed, we checked whether specific visual features (e.g., position, size, etc.) were represented in the recorded neuronal populations, by applying to the neuronal data a semi-supervised clustering method called the K-means clustering. This algorithm divides a dataset in clusters by assigning each point to the cluster with the nearest mean distance among the K possible alternatives (with the number K defined by the user).

We applied it to our dataset by first taking one of the stimulus properties and binning it in a meaningful way, and then by running the K-means algorithm over the neuronal data, setting K equal to the number of bins of the stimulus property. Once obtained the clusters from the K-means, we compared them to the categories obtained by binning the property under consideration; to compare to what extent the two different labelings (or groupings) of the stimuli overlapped, we computed the Adjusted Mutual Information, a well-known measure of accordance between clusters used in the machine learning community.

The Adjusted Mutual Information measures the accordance between two labelings of a set of objects by computing the mutual information between them. Given C and K as two distinct labelings of a set of objects, we can

define their entropy as

$$\begin{aligned} H(C) &= \sum_{i=1}^{|C|} P(i) \log_2(P(i)) \\ H(K) &= \sum_{i=1}^{|K|} P(i) \log_2(P(i)) \end{aligned} \tag{2.11}$$

where $P(i)$ is the probability of a random element from C or K falling into class C_i and K_i respectively. Mutual information is now written as

$$\mathrm{MI}(C, K) = \sum_{i=1}^{|C|} \sum_{j=1}^{|K|} P(i, j) \log_2 \left(\frac{P(i, j)}{P(i)P(j)} \right) \tag{2.12}$$

and can be normalized to the cross-entropy

$$\mathrm{NMI}(C, K) = \frac{\mathrm{MI}(C, K)}{\sqrt{H(C)H(K)}} \tag{2.13}$$

The Normalized Mutual Information unfortunately is not guaranteed to converge to 0 for a random assignment of the labels, especially in the case where the number of clusters is comparable to the number of objects. This problem can be circumvented by subtracting the expected value of the Mutual Information for a random assignment of the labels (Vinh et al., 2009) and renormalizing, obtaining the Adjusted Mutual Information (Vinh et al., 2010)

$$\mathrm{AMI} = \frac{\mathrm{MI} - E(\mathrm{MI})}{\max(H(C), H(K)) - E(\mathrm{MI})} \tag{2.14}$$

Both the clustering and the comparison were made using the scikit-learn Python library using Python3.5 (Pedregosa et al., 2011).

Chapter 3

Results

3.1 Information Theory

As a first step towards assessing what visual features are encoded by different visual areas, we computed the mutual information between single neuron response and several features, as detailed in Materials and Methods (section 2.4.3). Figure 3.1 **A** shows the amount of information about the position of the stimulus conveyed by neuronal response of well isolated units (single units or SUs), while Figure 3.1 **B** and **C** report the amount of information about the luminance of the stimulus, either weighted by the receptive field of each neuron (referred to as *RF luminance* in the following; panel **B**) or simply measured as the total luminance over the stimulus display (referred to as *total luminance* in the following; panel **C**) (see Tafazoli, Safaai, et al., 2017).

In the case of stimulus position (A) and total luminance (C), we observed a clear trend across V1-LM-LI-LL progression, with information about such low-level properties getting progressively discarded while moving up along the processing hierarchy. By contrast, information about RF luminance (B) did not show any significant trend, remaining relatively constant across all four visual areas. Figure 3.2 shows the same analyses applied to multi units (MUs), where the same trends observed for the SUs become even more apparent.

We applied a similar analysis to measure the amount of information that single neurons in the four areas carried about visual features that can be considered of either mid-level (such as orientation or aspect ratio of the stimuli; Figure 3.3) or high-level (such as object identity; Figure 3.4). Note that, in the latter case, mutual information between neuronal response and object identity is computed by taking into account all possible views (or appear-

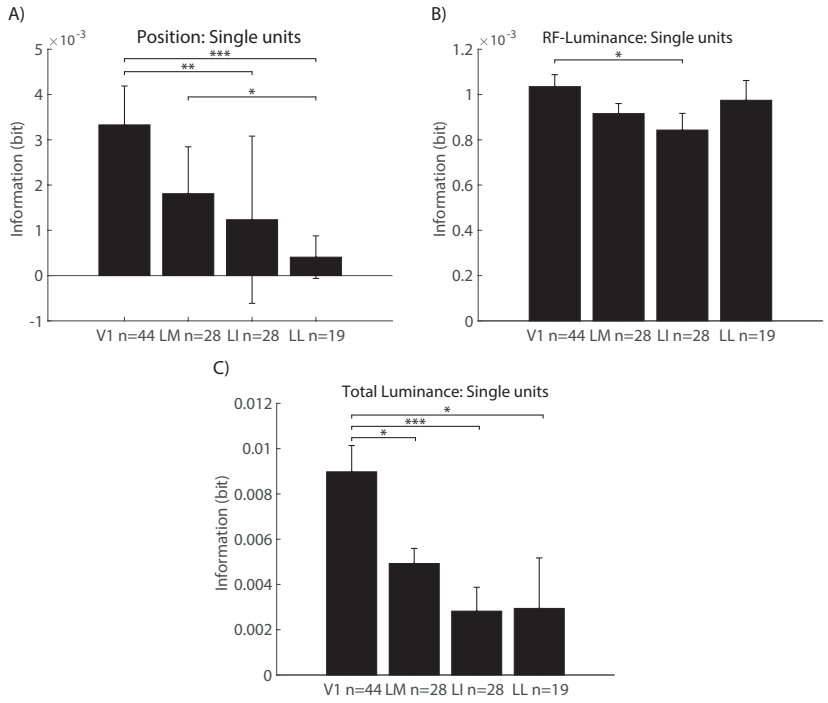

Figure 3.1. Mutual information between neural activity and stimulus features. Panel **A** shows the mutual information between neural response and the position of the stimulus on the X axis (single units). Panels **B** and **C** show instead the mutual information between the response and the luminance, computed only inside the receptive field or across the whole stimulus. Significant differences between areas are marked by the bars above them (1-tailed MannWhitney U test, Bonferroni-Holm corrected. $*p < 0.05$, $**p < 0.01$, $***p < 0.001$)

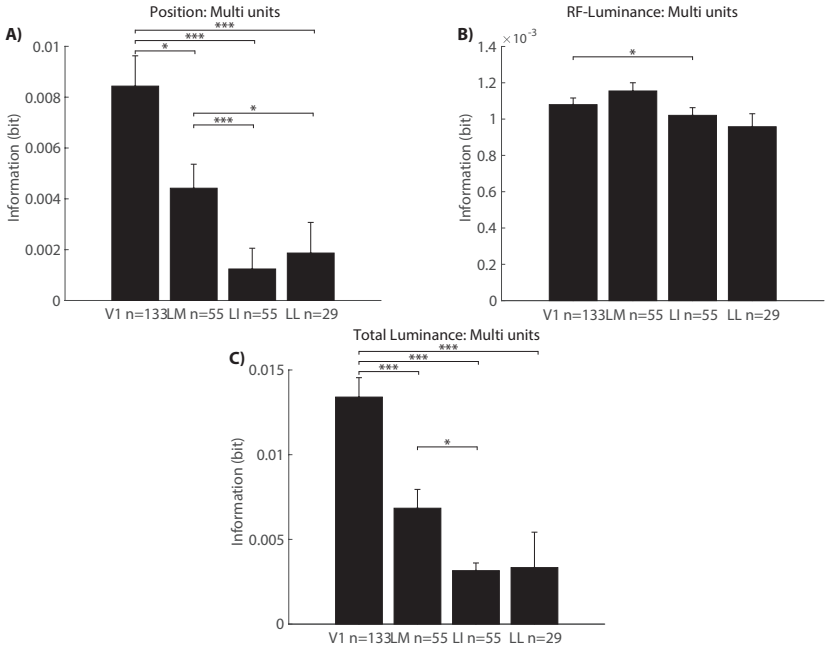

Figure 3.2. Mutual information between neural activity and stimulus features. Panel **A** shows the mutual information between neural response and the position of the stimulus on the X axis (multi units). Panels **B** and **C** show instead the mutual information between the response and the luminance, computed only inside the receptive field or across the whole stimulus (Median $\pm$ standard deviation. 1-tailed MannWhitney U test, Bonferroni-Holm corrected. $*p < 0.05$, $**p < 0.01$, $***p < 0.001$).

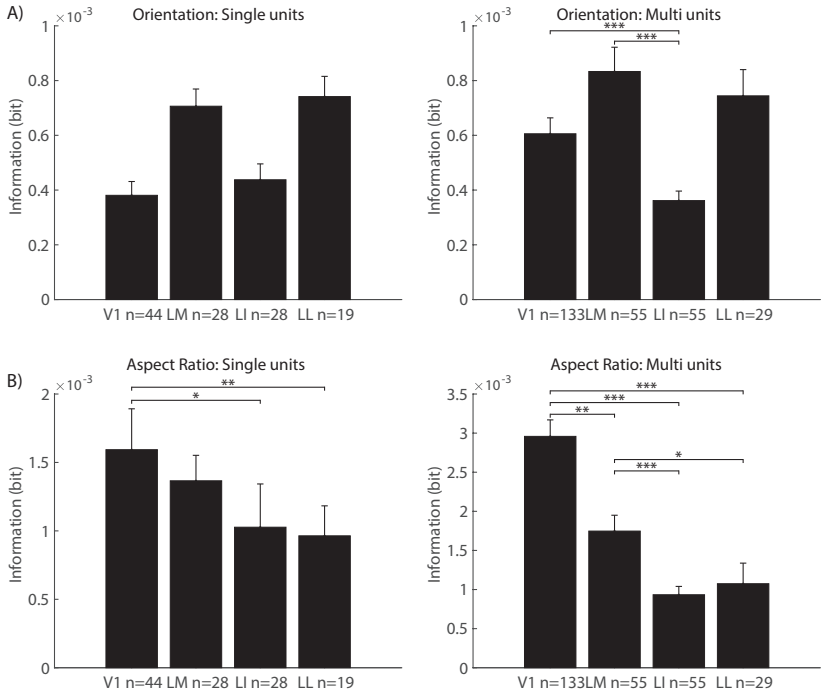

Figure 3.3. Mutual information between neural activity and mid-level stimulus features. Panel **A** shows the mutual information between neural response and the orientation of the stimulus' main axis for single and multi units. Panel **B** shows instead the mutual information between the response and the aspect ratio of the stimulus, again for single and multi units (1-tailed MannWhitney U test, Bonferroni-Holm corrected. $*p < 0.05$, $**p < 0.01$, $***p < 0.001$).

ances) of each object. Therefore, it is measure of the ability of single neurons to code object identity in a transformation-invariant way.

Information about orientation did not show any monotonic trend across the four areas, while information about aspect ratio significantly decreased from V1 to LL, following a trend that was similar to the one previously observed for the position and luminance information. Similarly, also the information about object identity decreased along the areas progression. This result may seem surprising, given that higher-order areas (such as LL) are expected to code more invariantly object identity, compared to low-level ones (such as V1). However, two considerations must be taken into account. The first is that, because of the data processing inequality, information can never increase along feed-forward processing hierarchy. This constraint does not strictly apply to the V1-LM-LI-LL progression, because the hierarchy is not strictly feed-forward (e.g., bypass routes exist that reach LL directly from V1) (Coogan and Burkhalter, 1993; Montero, 1993; Sanderson et al., 1991). In addition, the number of neurons per area becomes progressively smaller along the progression (Espinoza and Thomas, 1983) making it possible, in principle, for object information conveyed by single neurons to become larger from one area to the next, while preserving (or reducing) the total amount of information per area. More importantly, as shown by Tafazoli, Safaai, et al. (2017), a given object may be characterized by specific values of some low-level variable (e.g., luminance) in spite of the view changes it undergoes. For instance, a black object will tend to have lower luminance values than a bright object, regardless of size, position or pose variation. As such, a given low-level property may act as a cue for object identity, thus acting as a confound in the assessment of the amount of invariant object information conveyed by single neurons.

A way to address this confound is to compute conditional mutual information, i.e., to measure the information carried about a given (mid- or high-level feature), while fixing another (low-level) feature (see Materials and Methods 2.4.3). Figures 3.5 and 3.6 show this analysis applied to orientation and object identity, while fixing position or luminance.

For instance, the total amount of joint information carried by single neurons about orientation and position (full bars in Figure 3.5 **A**) decreased along the areas progression. However, the information about orientation only (in fixed position bins) remained constant (blue portion of the bars), both in the case of SUs and MUs. As a result, the fraction of information about orientation only (i.e., the ratio between conditional and joint mutual information) sharply increased from V1 to LL. This indicates that single neurons along the progression become capable of signaling the orientation of the objects, in a way that is increasingly less polluted by confounding low-level

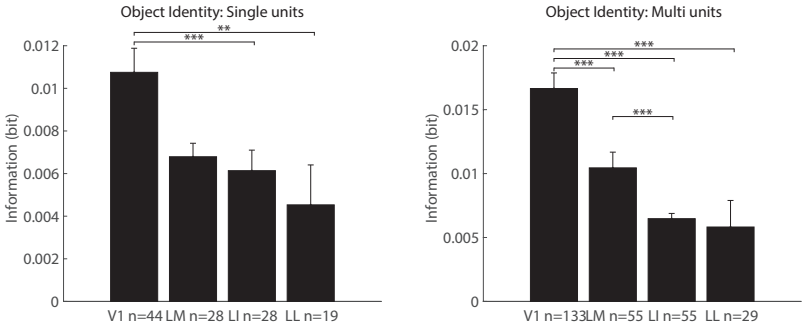

Figure 3.4. Mutual information between neural activity and stimulus identity for single and multi units. Stimulus identity is defined as the identity of the object represented in it, so that stimuli sharing the same id will be different views of the same object (1-tailed MannWhitney U test, Bonferroni-Holm corrected. $*p < 0.05$, $**p < 0.01$, $***p < 0.001$).

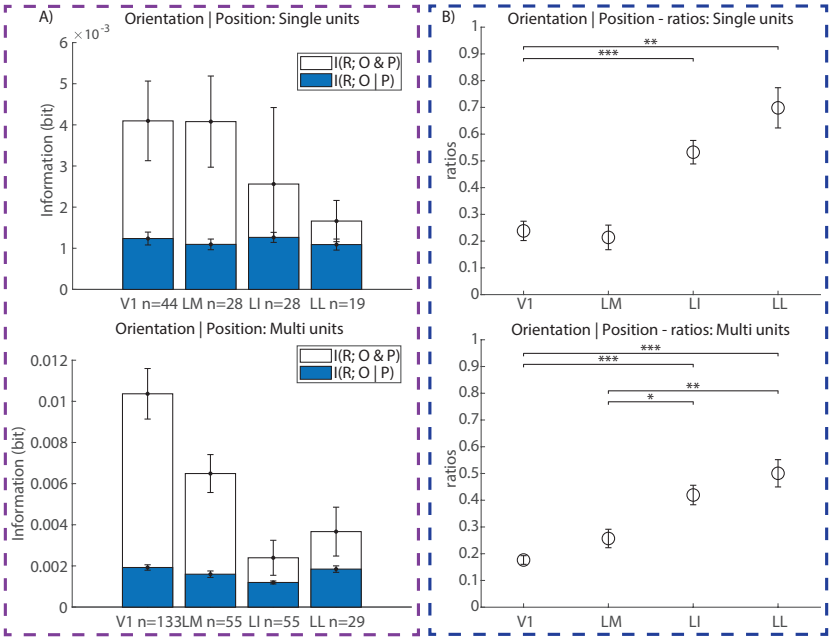

Figure 3.5. Conditional mutual information between neural activity and stimulus orientation, fixing position. In panel **A** are reported the conditional information about orientation ($I(R;O|P)$, blue bar) and the total information about position and orientation ($I(R;O\&P)$, empty contour) for single and multi units. Panel **B** shows the ratio between conditional and total mutual information in the four different areas, for single and multi units (1-tailed MannWhitney U test, Bonferroni-Holm corrected. $*p < 0.05$, $**p < 0.01$, $***p < 0.001$).

information. In other words, orientation becomes represented in a way that is increasingly more explicit, while information about position is gradually discarded. This trend is evident both at the level of single and multi unit activity.

Figure 3.6 shows that a similar trend (although with a shallower slope) was also found when we computed the conditional mutual information between object identity and firing rate, having fixed either object position or luminance. The increase in the fraction of object information conveyed by the neuronal responses, relative to the total information carried about both object identity and either position or luminance, is particularly prominent in MUs but is also present, although weaker in magnitude, in the SUs, when conditioning over position. This implies that, although in absolute terms the invariant object information decreases along the areas progression (blue portion of the bars), the fraction of coding power that each neuron allocates to represent object identity increases from V1 to LL. As such, the information about object identity becomes more explicit along the pathway and, potentially, easier to read out from downstream neurons.

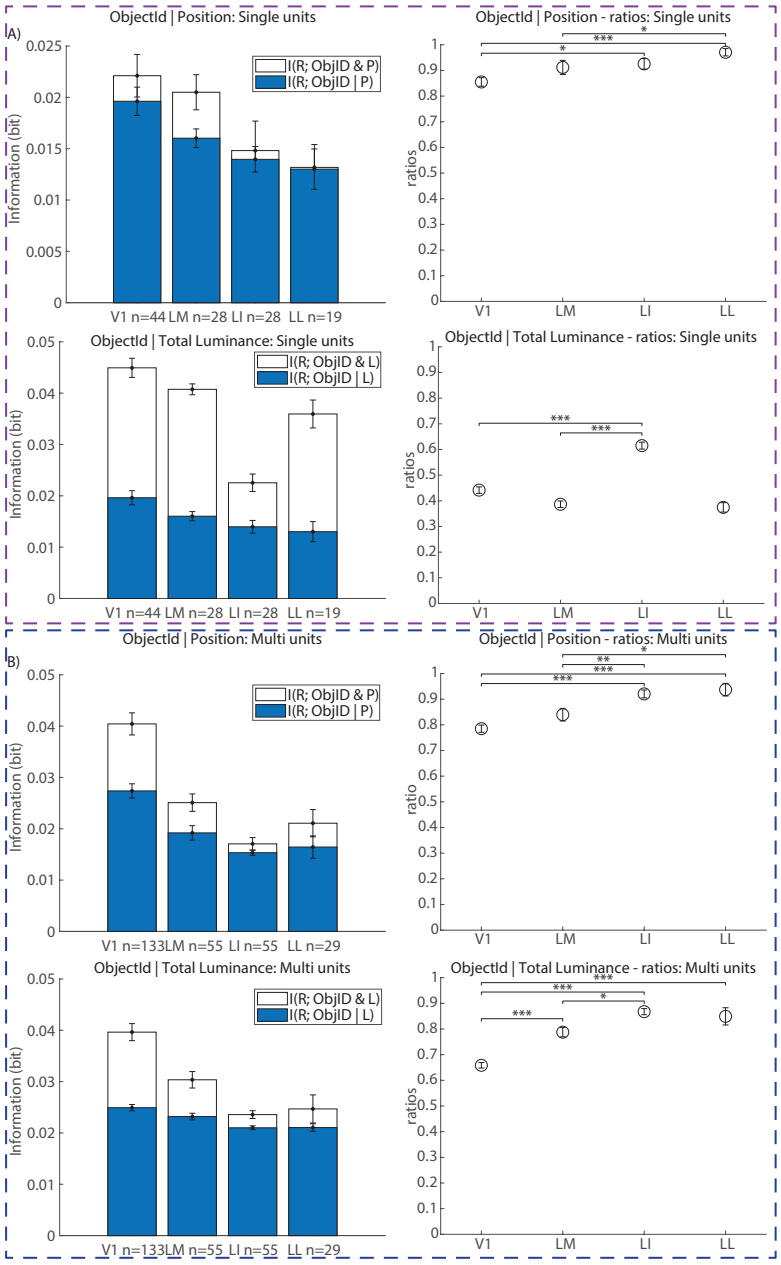

Figure 3.6. Conditional mutual information between neural activity and stimulus identity, fixing position and luminance. Both the conditional mutual information and its ratio with the total information are reported, as in figure 3.5. Panel **A** shows the results for the single units, while panel **B** for the multi units (1-tailed MannWhitney U test, Bonferroni-Holm corrected. $*p < 0.05$, $**p < 0.01$, $***p < 0.001$).

3.2 Representational Dissimilarity Analysis

It is well established that neurons encode information collectively to drive perceptual decision. Therefore, once assessed the tuning properties of rat visual neurons at the single-cells level (see previous section), it was essential to investigate how neuronal populations encode specific kinds of visual information along the V1-LM-LI-LL progression. This analysis can be carried out using two alternative (largely complementary) classes of approaches: supervised or unsupervised pattern classification tools. While the former have been extensively applied in both monkey studies of the ventral stream (Baldassi et al., 2013; DiCarlo, Zoccolan, et al., 2012; Hung et al., 2005; Rust and DiCarlo, 2012) and recent investigations of rat visual cortical areas (Tafazoli, Safaai, et al., 2017; Vermaercke, Gerich, et al., 2014), they have the disadvantage of imposing a given classification task over the collected neuronal data. By contrast, unsupervised approaches, such as clustering methods, allow patterns that are present in the data to naturally emerge, thus possibly revealing a richer structure in the neuronal representation of visual stimuli (Baldassi et al., 2013; Kiani et al., 2007; Kriegeskorte, Mur, et al., 2008). The first unsupervised population analysis we performed was based on the computing the Representational Dissimilarity Matrix (RDM). As explained in section 2.4.4 of the Material and Methods, we built the pairwise distance matrix by computing, for every possible pair of object conditions i and j, the correlation distance $(1-\rho(\nu_i, \nu_j))$, where the correlation coefficient rho is computed over the population response vectors vi and vj evoked by the two stimuli. The correlation distances can be visualized as a color map, which can it turn be visually inspected to look for possible structures (i.e., blocks of highly correlated or poorly correlated stimuli). Obviously, the emergence of such blocks will depend on the way the data (i.e., the stimuli) are arranged along the matrix.

In figure 3.7 we report the representational dissimilarity matrices obtained from single units (panels **A** and **B**) and from multi units (panels **C** and **D**) using two different methods to reorder the matrix. In panels **A** and **C** we used the result of a hierarchical binary cluster tree, computed using the correlation distance and the complete linkage method (furthest distance), while in panels **B** and **D** we sorted the stimuli according to the position on the X axis.

By construction, arranging the data according to the results of the hierarchical clustering algorithm, result in several blocks of stimuli with more or less homogeneous correlation (distance). In fact, the hierarchical clustering is based on the distances values reported on the RDM. The usefulness of plotting the data according to this ranking is to compare the intensity and size

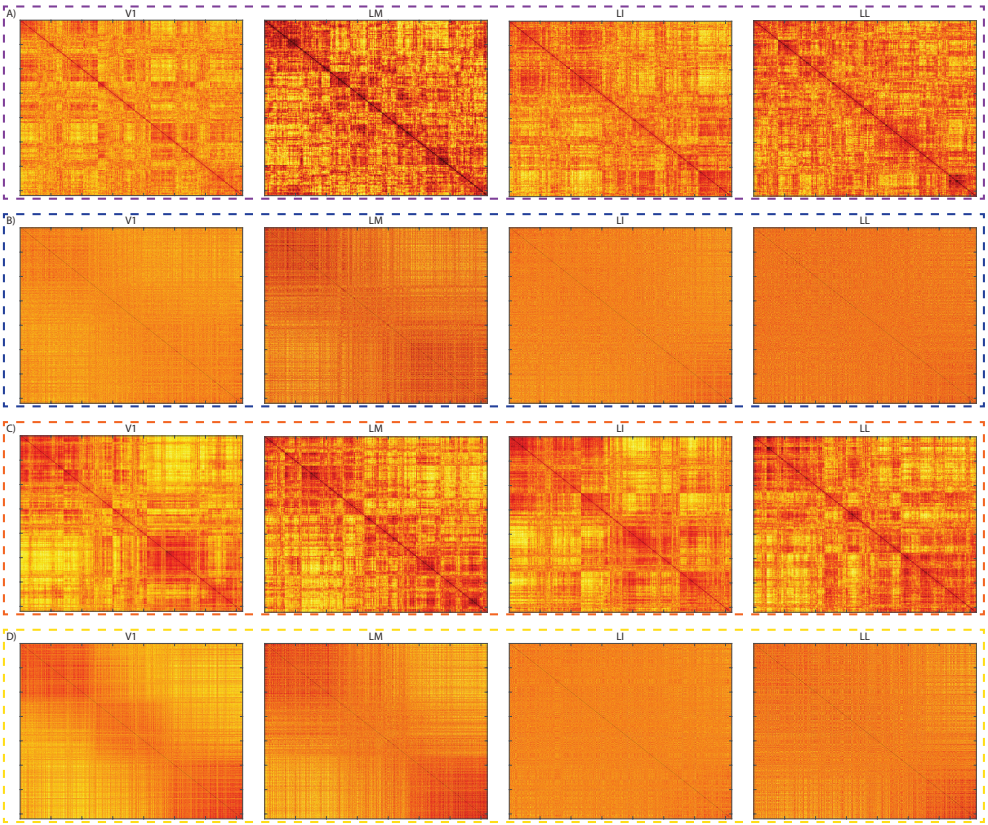

Figure 3.7. Representational dissimilarity matrices for single and multi units. In panels **A** and **C** are the RDMs (for single and multi units, respectively) ordered according to complete linkage. In panels **B** and **D** are the RDMs built by ranking the stimuli according to their position.

of the resulting blocks across the areas. This gives an idea of how compactly the data are clustered in the neuronal representational spaces provided by the four areas. As shown in Figure 3.7 **A** and **C**, the block-like structures are more evident in the case of the multiunit activity (panel **C**), where they tend to become sharper and more evident as we proceed from V1 to LL. Overall, this can be taken as a qualitative indication that visual objects better map into a categorical representation in LL than V1. Another way to arrange the axes of the RDM is by sorting the stimuli according to specific stimulus parameters. This can be especially useful to assess whether variations of low-level visual properties are represented in the recorded populations. Figure 3.7 **B** and **D** show the result of this analysis, when the object conditions were arranged along the axes of the matrix according to their position. In V1, the dissimilarity matrices clearly show three blocks of similar stimuli (again, better visible in the multiunit activity; panels **D**), corresponding to the three main positions where the stimuli were presented (see Materials and Methods 2.1.2). This block diagonal structure becomes more blurred in LM, and disappears completely in LI and LL. Therefore, the RDM analysis shows a better ability of the neuronal populations in low-level areas to represent low-level features, such as object position, as compared to higher-level areas, thus confirming the result of the mutual information analysis applied to single cells.

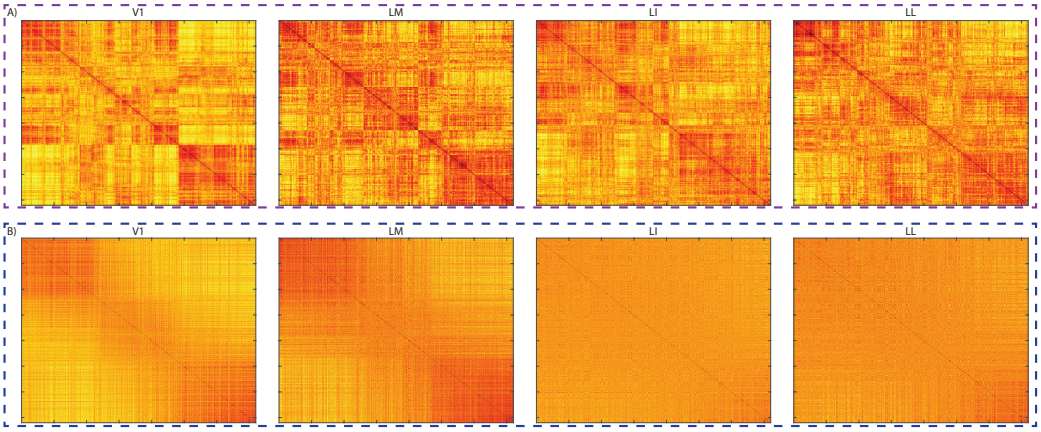

Figure 3.8. Representational dissimilarity matrices obtained joining single and multi units. The trends emerging separately in well isolated single units and in multi units are present even when they are joined.

Since the results of the RDM analysis in well isolated single units shows very similar results with the one performed with multi units (as already seen for the mutual information analysis presented in the previous section), we repeated the analysis pooling together the two sets of neuronal data, so as to

increase the statistical power of our test. The results are shown in figure 3.8 and confirm the trend observed when the two sets of units were considered separately. For this reason, in the following analyses, we always pooled the single and multi units together.

3.3 Quantification of visual features along the hierarchical clusters

The RDM provides a graphical intuition of the kind of information that can be represented by a neuronal population, but such intuition needs to be quantified by further analyses. To check whether position information was indeed better represented by low-level areas V1 and LM, as compared to higher-order areas LI and LL, we measured the variance of stimulus position along the depth of the tree produced by hierarchical clustering (see Materials and Methods 2.4.4).

As shown in Figure 3.9, since the small clusters (nodes) near the bottom of the tree (left-hand side of the abscissa) contain few elements, the variance can fluctuate widely, resulting in very noisy curves. For this reason we lowpass filtered the curves, using a moving average with a 64 samples width and an 8 samples stride.

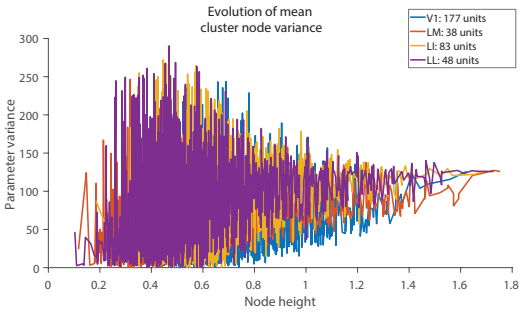

Figure 3.9. Evolution of the variance of stimulus position on the X axis. The small size of the clusters near tree height 0 cause the variance to have ample fluctuations, hence the need to smooth the curve.

Here, it can be appreciated how the curve relative to V1 (blue) is well below all the other curves, followed by LM (red) and then by LI and LL (yellow and purple). That is, at every depth along the hierarchical tree, the objects under a node have a more similar position (i.e., smaller position variance) in V1 than in all other areas, with LM displaying a similar behavior, relative to LI and LL. This suggests that the two lowest-level areas (V1

and LM) map the stimuli that appear in neighboring positions in the visual field near each other in their representation space, while in LI and LL such retinotopic representation is much weaker. A similar result holds for another low-level property: luminosity (Figure 3.10, right) although, in this case, the separation between LM, LI and LL vanishes, leaving only V1 well isolated from the other areas. The separation vanished completely when we analyzed the orientation (figure 3.11). Overall, these plots suggest that lower level areas (V1 and LM) encode predominantly low level features like position and luminosity.

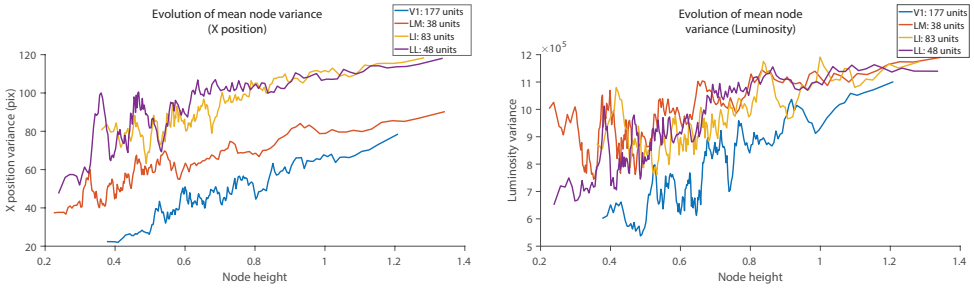

Figure 3.10. Evolution of parameter variance across a hierarchical tree (Position and Luminosity).

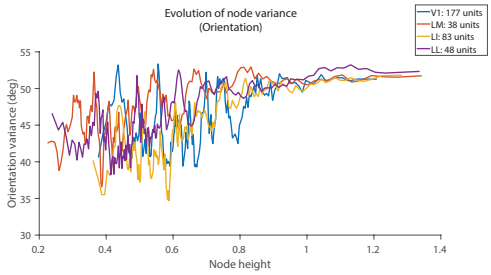

Figure 3.11. Evolution of parameter variance across a hierarchical tree (Orientation).

3.4 Dimensionality Reduction Analysis

We further explored what kind of visual information neurons encoded in each visual area, by applying another unsupervised multivariate analysis: Principal Components Analysis (PCA). PCA was performed on the activity (i.e., the population vectors) of the four areas. The rationale of this analysis was to check whether any principal component existed that could be associated

to the variation of some visual property shared by the objects of our stimulus set. To inspect whether this was the case, we considered various visual features and, for each of them, we computed its correlation with each principal component (PC). We then projected each stimulus on the 2-dimensional space defined by the two principal components that had the highest correlation with the visual property under examination, and we color coded the value of such property over the stimulus set. The resulting scatter plots allowed visually assessing whether there was a systematic variation of the visual property over one or more of the principal components.

As shown in Figure 3.12 (left), in V1, we found a sharp gradient of stimulus position along one of the first principal components: PC2, which, in fact, was highly correlated with position (first bar in the histogram). In the other areas we also found that some of the principal components correlated with stimulus position, but such correlation was strong only in LM, while it was much weaker in LI and LL. As a result, the separation of the stimuli according to position became progressively more fuzzy along the areas progression. In addition, we found that the principal component that mostly correlated with position was higher in LI and LL (respectively, PC5 and PC3) than in V1 and LM (PC2). This indicates that, beside the lower magnitude of the correlation, the PC that accounted for the variation in the position of the stimuli, explained less variance in LI and LL, as compared to V1 and LM.

By comparison, a quite different trend was found for luminosity (Figure 3.12, right), since the magnitude of the correlation did not vary much across the areas. In addition, in V1, the principal component that was mostly correlated with luminance was surprisingly high (PC15), while it was surprisingly low in LL (PC2). As hypothesized in section 2.4.3, the latter finding could be explained by the fact that LL could code features of higher level that cooccur with luminosity and that cannot be told apart by this kind of analysis. With regard to V1, the high rank of the mostly correlated PC might stem from the fact that lower order principal components already code for other features (i.e., position), so luminosity is encoded by subsequent components. Unfortunately we still need to gather more data and refine the analysis to investigate the issue further and verify these hypotheses.

When we applied PCA to mid-level features, as shown in Figure 3.13 for stimulus orientation, the structure found in the previous analysis is completely lost, and even the mostly correlated principal components show little correlation with orientation. Although this does not conflict with the results coming from the information theory analysis (see in particular figure 3.3, top row), it might be showing a limitation of the PCA in its ability to capture complex dependencies between neural activity and stimulus features. These results confirm nonetheless that the activity of low level areas better encodes

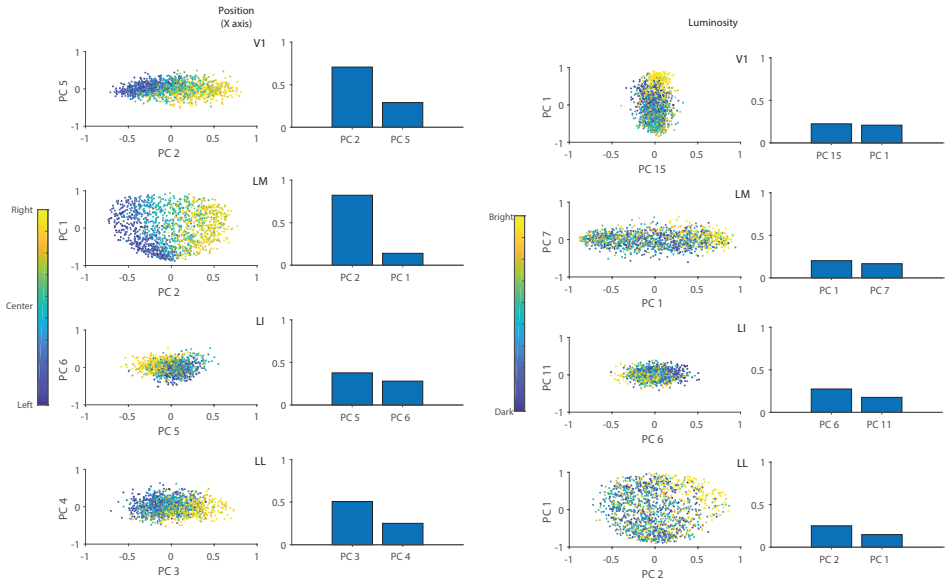

Figure 3.12. Principal Component Analysis for Position and Luminosity. On the left are the scatter plots of the two principal components that correlate most with position or luminosity. The value of the parameter for each stimulus is color coded in its dot. On the right are the absolute values of the correlation between the principal component and luminosity or position.

low-level features, unlike the activity from higher-level areas.

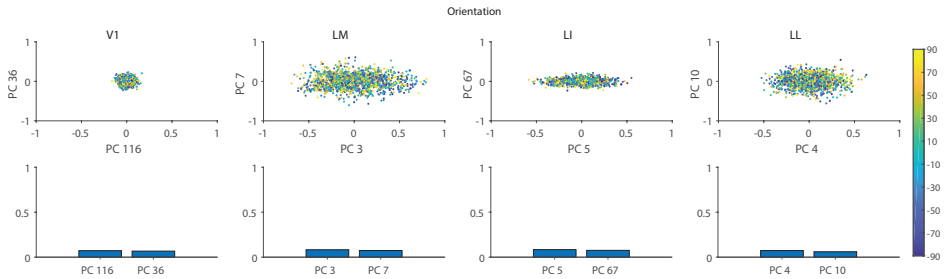

Figure 3.13. Principal Component Analysis for Orientation. On the top are the scatter plots of the two principal components that correlate most with orientation. The value of the parameter for each stimulus is color coded in its dot. On the bottom are the absolute values of the correlation between the principal component and the orientation of the stimulus.

3.5 K-means Clustering Analysis

As a final step in our investigation of how our visual objects were mapped in the neuronal representational space, we partitioned the stimuli in discrete clusters by applying a k-means algorithm to the set of population vectors. The rationale was to compare the clustering produced by the k-means with the partitions of the stimuli obtained by binning various stimulus properties (e.g., position or luminance). To make this comparison straightforward, the number of k-means clusters was set to the number of bins in which the stimuli were divided according to the given visual property under consideration. The neuronal-based clustering and the feature-based binning were then compared by computing the adjusted mutual information (as described in Materials and Methods 2.4.4), a metric that measures to what extent two different partitions of the same data set match.

Figure 3.14 shows the adjusted mutual information between neuronal-based (k-means) clustering and partitions of the stimulus set that were obtained by binning the visual objects according to three different features: position, luminosity, and area. In case of position the stimuli were divided in three bins (categories), corresponding to the three main presentation positions(left, center and right) used during the presentation protocol (see Materials and Methods 2.4.4). For luminosity and the area, we divided the stimuli in 10 categories and ran the k-means accordingly.

These analyses revealed that the agreement between neuronal-based clustering and feature-based binning was greater in low-level areas (V1 and LM)

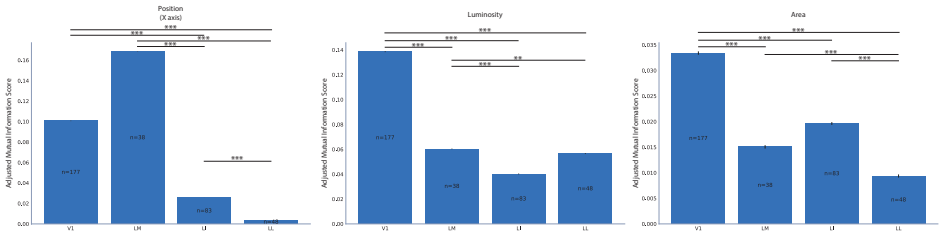

Figure 3.14. Comparison between feature defined clusters and K-means clusters for low level features. We report here the results for clusters based on X position, luminosity, and area (left to right, respectively). Mean ± SEM, significant differences between areas are marked by the bars above them (1-tailed MannWhitney U test, Bonferroni-Holm corrected. $*p < 0.05$, $**p < 0.01$, $***p < 0.001$)

than in high level ones (LI and LL); this was especially evident in the case of position, where clusters in LI and LL bear almost no information about stimulus position. For area and luminosity the decrease along the extrastriate areas (LM, LI, and LL) is less monotonic, but we still find a stark difference between V1 and all three extrastriate areas.

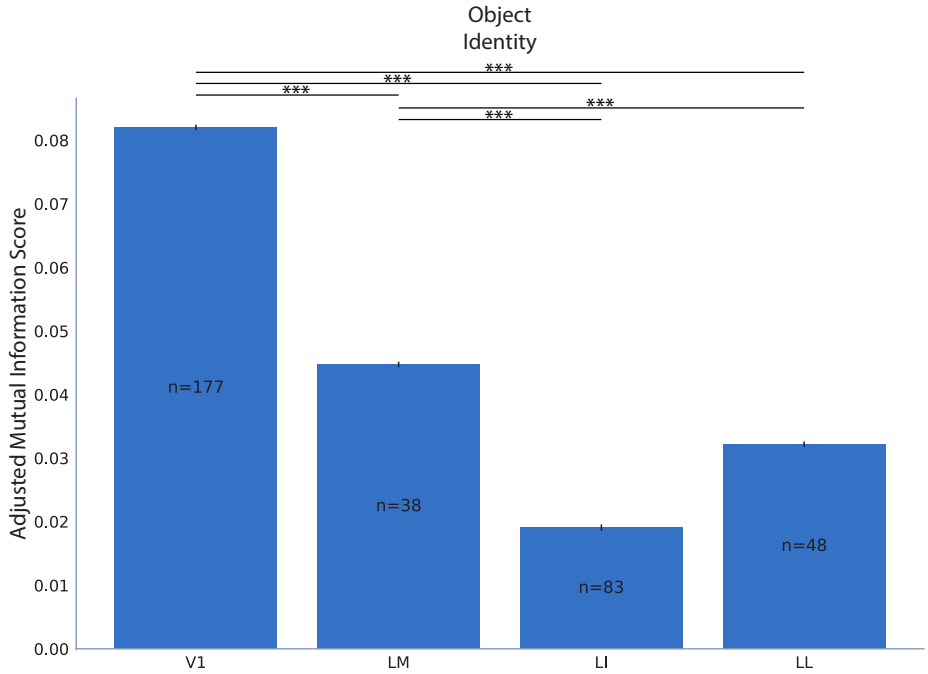

Figure 3.15. Comparison between K-means clusters and object ID. (1-tailed MannWhitney U test, Bonferroni-Holm corrected. $*p < 0.05$, $**p < 0.01$, $***p < 0.001$)

Next, we investigated to what extent the four areas were able to tell the

different objects apart, in spite of their variations in terms of pose, position, orientation, etc. To this aim, we ran the K-means clustering with 40 clusters, one for each object in the stimulus set, and we compared the result with the actual object identity label associated to each object view (this amounts to partition the stimulus set in 40 categories with 36 elements each). The trend obtained across the cortical pathway (Figure 3.15) showed that primary visual cortex was the area that better supported the segregation of the stimuli according to object identity. This was surprising, given that a low-level area such as V1 should not be able to code the identity if visual objects in way that is invariant to transformations, at lest not better than downstream areas. This is certainly the case of monkey ventral stream, but a similar behavior has been also for he V1-LM-LI-LL progression in the rat brain (Tafazoli, Safaai, et al., 2017). However, as already shown in the information theoretic analysis applied to single cell responses, this result could be due to some interaction between object identity and lower-order visual properties. Specifically, we suspected that stimulus luminosity was playing a role in determining the trend of Figure 3.15, since, across the views of any given object, luminosity did not vary much (e.g., on average, across views, dark objects tended to remain dimmer than light objects). To address this possible confound, we proceeded as previously done for the mutual information analysis applied to single-cell data (see Section 3.1). That is, we assessed matching between neuronal-based clustering and partitioning of the stimuli according to object identity, but restricting this comparison only to a subset of the stimuli, as obtained by fixing some of the low-level visual properties.

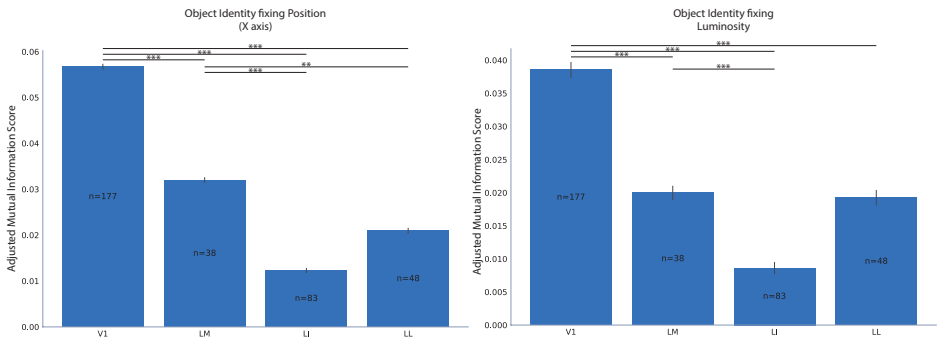

Figure 3.16. Comparison between K-means clusters and object ID when fixing the values of low-level properties. The plot on the left shows the mutual information computed at fixed X position, while the plot on the right the mutual information computed fixing luminosity. (1-tailed MannWhitney U test, Bonferroni-Holm corrected. $*p < 0.05$, $**p < 0.01$, $***p < 0.001$)

We started by fixing separately position and luminosity (figure 3.16 left

and right, respectively). We divided luminosity in 10 bins and for each bin we repeated the analysis of figure 3.15, taking into account only those objects with at least 5 views falling into that bin. This yielded 10 different histograms, each showing the matching between neuronal-based clustering and partitioning of the stimuli according to object identity in each luminance bin. These histograms were then averaged together, to yield the average trend shown in Figure 3.16, right. We performed the same operation for the position, which we divided in three bins (left, center, right), obtaining the trend shown in Figure 3.16, left. However, these trends were not different from those already obtained in Figure 3.15: V1 remained the area that better supported the segregation of the objects according to their identity, although there was, in both plots, a reduction in the magnitude of the adjusted mutual information for all the four areas.

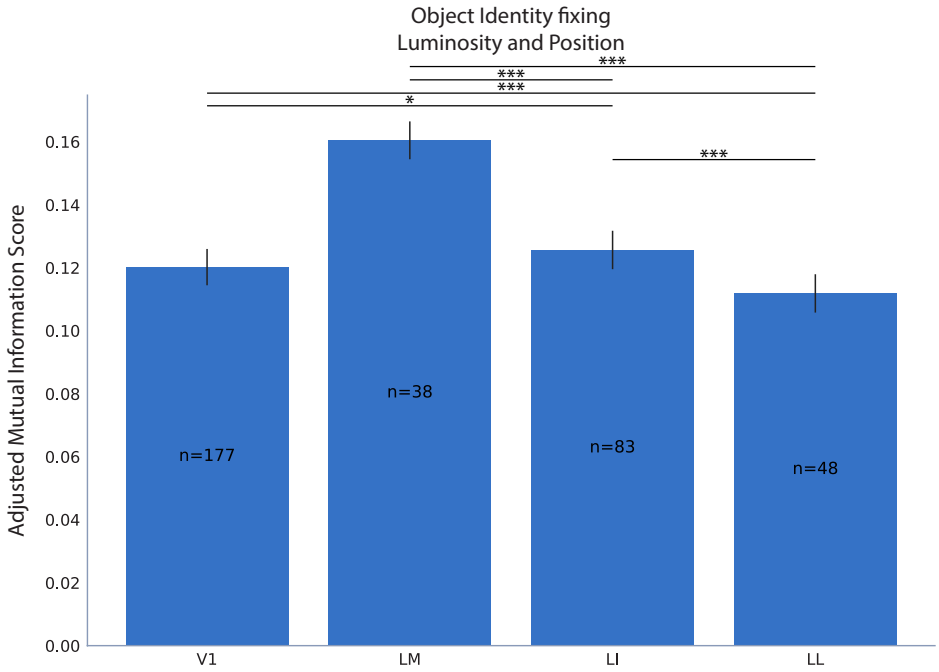

Figure 3.17. Comparison between K-means clusters and object ID when fixing both X position and luminosity. (1-tailed MannWhitney U test, Bonferroni-Holm corrected. $*p < 0.05$, $**p < 0.01$, $***p < 0.001$)

We thus decided to go further in our approach of fixing low-level visual properties, before performing the k-means clustering analysis. Specifically, we fixed both position and luminosity at the same time, using the same binning we applied before, when they had been fixed separately. The resulting trend (Figure 3.17) showed that the four areas were comparably good at

representing object identity, in a way similar to what happened when we computed mutual information between neural response and object identity for single neurons (figure 3.6), after fixing low-level properties. Overall, these findings confirmed that the true ability of neuronal populations to support invariant discrimination of object identity is preserved along the areas progressions; with the advantage that, in higher-order areas (i.e., LI and LL), object information is not polluted with low-level features, which can make it easier to read object identity from these representations.

3.6 Dimensionality Estimation Across Areas

The last analysis we report is the estimation of the intrinsic dimensionality of the manifolds where our object conditions lay in the representational spaces provided by the four areas. Unfortunately, when merging SUs and MUs, the block analysis (explained in Materials and Methods; section 2.4.4) did not show a plateau for any of the areas (figure 3.18), making the estimate of the intrinsic dimensionality of our dataset unreliable. For this reason we restricted the analysis to MUs only.

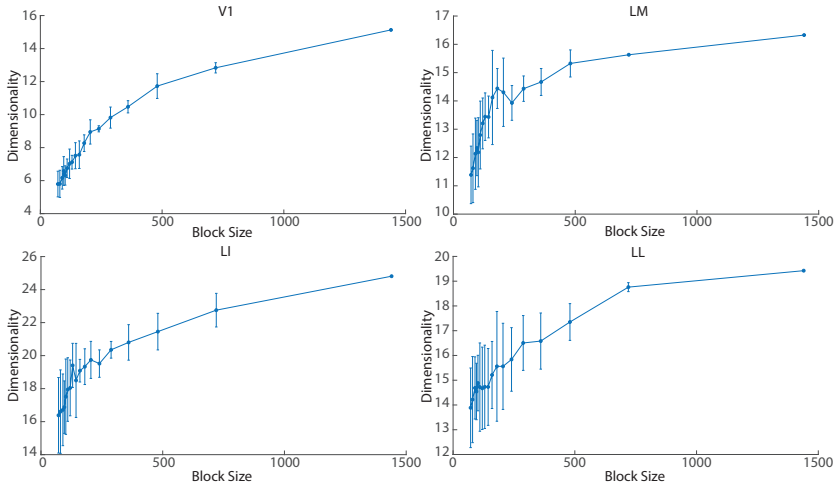

Figure 3.18. Block analysis result for single and multi units pooled together. Note the absence of a plateau in any of the areas.

As a first test of the goodness of our analysis, we checked that the plot of the ratio of the first and second neighbor distances $(\log(\nu), -\log(1 - F(\nu))$ (see Materials and Methods) followed a straight line, and, therefore, did not present evidence of multiple trends, which would have been an index of the presence of several manifolds of different dimensions upon which the data

were lying. In all the visual areas, no evidence of such multiple trends was found (as shown, for area LM, in Figure 3.19).

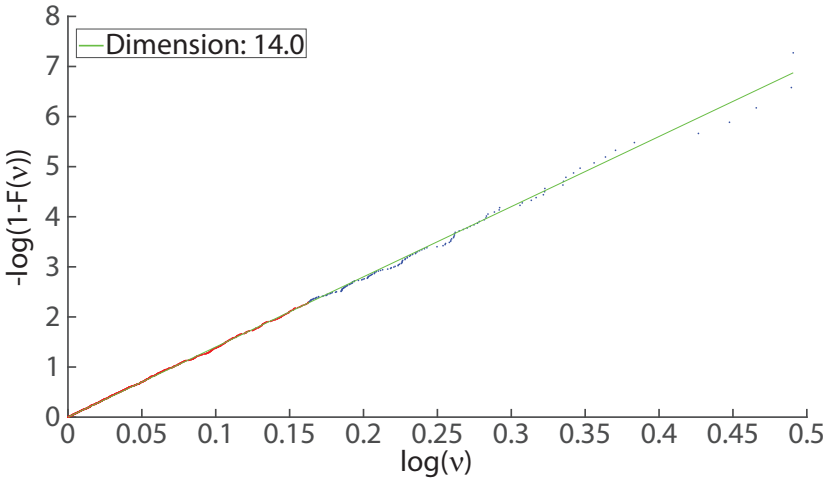

Figure 3.19. Plot of the ratios of the first and second neighbor distances. The values of ν are well behaved and follow a straight line, whose coefficient indicates the intrinsic dimension is 14 (green line). The fit is only computed on 90% of the data (indicated in red).

We then applied again the block analysis and we found that, with the exception of V1, a plateau could be found in all visual areas (Figure 3.20). These plateaus indicate that the intrinsic dimensionality of the neural representation of our dataset are 13.2 for LM, 19 for LI, and 14.5 for LL.

These values suggest that the intrinsic dimensionality of object representations increases at first when moving towards higher level areas, only to decrease again past area LI. These conclusions, however, are still preliminary, since we need to increase our data sample and we need to understand what trends artificial convolutional neuronal networks would display with regard to intrinsic dimensionality.

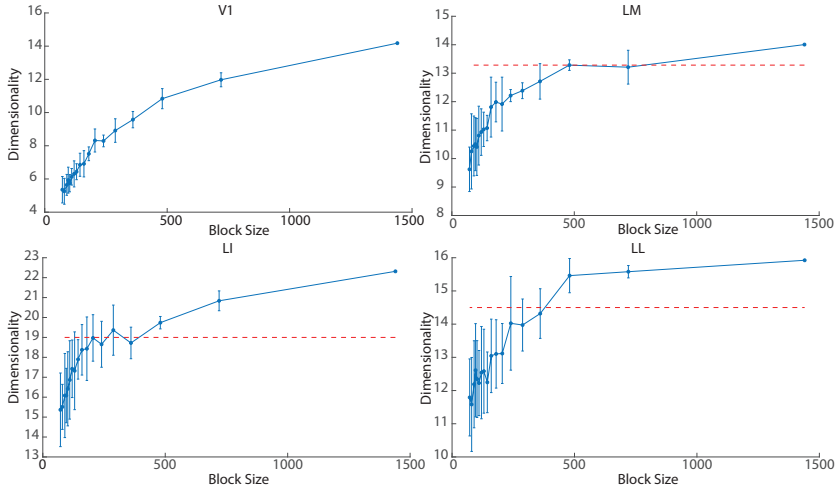

Figure 3.20. Result of the block analysis for multi units only. Where present, plateaus have been marked by a red dashed line at its dimensionality.

Chapter 4

Discussion

Overall our study provides more evidence in favor of the existence of high-level cortical processing of visual object information in rats, and of the hierarchical organization of rats' visual cortical areas. The analysis of neuronal responses using information theory confirms the progressive reduction of information about low-level features reported by Tafazoli, Safaai, et al. (2017), and this trend is also confirmed at the population level by the comparison between response-based clusters and feature-based categories, by the analysis of the relations between principal components and stimulus properties, and by the analysis of the variability of such properties along a hierarchical clustering tree.

The representational dissimilarity analysis (Kriegeskorte, Mur, et al., 2008) on the other hand shows the presence of a structure in the representation of the stimulus set we used, signaling that stimuli are categorized according to some feature or conjunction thereof. This structure is partially explained by the coding of low-level properties (specifically position along the x axis), but only in low-level areas of the visual hierarchy. This suggests that the structure found in the higher-level representations is formed using a combination of several low- and mid-level features which are yet to be determined.

Interestingly, during our analyses we found rather consistent results both when analyzing well-isolated single units and when analyzing multi-units, a result also confirmed in the population analyses. This prompted us to merge well-isolated single units and multi-units, at which point we once again confirmed our results. Thus, we concluded that small local subpopulations of neurons, such as those recorded as multi-units, effectively behaved like single neurons. This result is in accordance with the mean-field approximation, which theorizes a consistent dynamics for local neural populations under stationary conditions such as the ones provided by deep anesthesia (Amit

and Brunel, 1993; Mattia and Sanchez-Vives, 2012).

As future perspectives we are first planning to increase the number of recorded neurons; although we managed to obtain interesting results from the data in our hands, the low number of recorded neurons still remains a limitation, if not the main limitation, of our study. We are also planning to conduct in parallel other experiments on awake, head-fixed rats. This change of paradigm will allow us both to record more active units and in greater number, and to compare the analyses coming from anesthetized animals with the ones coming from awake animals, thus highlighting the potential effects of deep anesthesia on high-level processing. Finally, we are also going to explore, in a parallel project that started last spring, how the processing of visual stimuli is modified in rats raised in visually controlled environments. Specifically, we will test the role of early postnatal exposure to the natural spatio-temporal statistics of the visual environment in determining the processing properties of rat visual cortex. This will be possible by rearing newborn rats either in natural visual environments, i.e., exposed to natural movies, where the temporal continuity of the visual input is preserved, or in altered environments, i.e., exposed to movies, whose frames have been randomly shuffled, thus breaking the temporal continuity of the visual input.

Overall, the results obtained in our study, along with earlier findings about rat visual behavior (Alemi-Neissi et al., 2013; De Keyser et al., 2015; Rosselli et al., 2015; Tafazoli, Di Filippo, et al., 2012; Vermaercke and Beeck, 2012; Vinken, Vermaercke, et al., 2014; Zoccolan, 2015; Zoccolan et al., 2009) and cortical processing abilities (Tafazoli, Safaai, et al., 2017; Vermaercke, Gerich, et al., 2014; Vinken, Van den Bergh, et al., 2016; Vinken, Vogels, et al., 2017), support the use of this species as an interesting model to investigate the neuronal basis of visual object recognition.

Bibliography

Alemi-Neissi, A., Rosselli, F. B., and Zoccolan, D. (2013), "Multifeatural shape processing in rats engaged in invariant visual object recognition", *Journal of Neuroscience*, 33 (14): 5939–56.

Amit, D. J. and Brunel, N. (1993), "Adequate input for learning in attractor neural networks", *Network: Computation in Neural Systems*, 4 (2): 177–94.

Baldassi, C., Alemi-Neissi, A., Pagan, M., DiCarlo, J. J., Zecchina, R., and Zoccolan, D. (2013), "Shape similarity, better than semantic membership, accounts for the structure of visual object representations in a population of monkey inferotemporal neurons", *PLoS computational biology*, 9 (8): e1003167.

Berardi, N., Pizzorusso, T., and Maffei, L. (2000), "Critical periods during sensory development", *Current opinion in neurobiology*, 10 (1): 138–45.

Borst, A. and Theunissen, F. E. (1999), "Information theory and neural coding", *Nature neuroscience*, 2 (11): 947–57.

Carandini, M. and Churchland, A. K. (2013), "Probing perceptual decisions in rodents", *Nature neuroscience*, 16 (7): 824–31.

Coogan, T. A. and Burkhalter, A. (1993), "Hierarchical organization of areas in rat visual cortex", *Journal of Neuroscience*, 13 (9): 3749–72.

De Keyser, R., Bossens, C., Kubilius, J., and Beeck, H. P. O. de (2015), "Cue-invariant shape recognition in rats as tested with second-order contours", *Journal of vision*, 15 (15): 14–4.

Desimone, R., Albright, T. D., Gross, C. G., and Bruce, C. (1984), "Stimulus-selective properties of inferior temporal neurons in the macaque", *Journal of Neuroscience*, 4 (8): 2051–62.

DiCarlo, J. J. and Cox, D. D. (2007), "Untangling invariant object recognition", *Trends in cognitive sciences*, 11 (8): 333–41.

DiCarlo, J. J., Zoccolan, D., and Rust, N. C. (2012), "How does the brain solve visual object recognition?", *Neuron*, 73 (3): 415–34.

Espinoza, S. G. and Thomas, H. C. (1983), "Retinotopic organization of striate and extrastriate visual cortex in the hooded rat", *Brain research*, 272 (1): 137–44.

Fabre-Thorpe, M., Richard, G., and Thorpe, S. J. (1998), "Rapid categorization of natural images by rhesus monkeys", *Neuroreport*, 9 (2): 303–8.

Facco, E., d'Errico, M., Rodriguez, A., and Laio, A. (2017), "Estimating the intrinsic dimension of datasets by a minimal neighborhood information", *Scientific reports*, 7 (1): 12140.

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

Földiák, P., Xiao, D., Keysers, C., Edwards, R., and Perrett, D. I. (2004), "Rapid serial visual presentation for the determination of n eural selectivity in area STSa", *Progress in brain research*, 144: 107–16.

Gavornik, J. P. and Bear, M. F. (2014), "Higher brain functions served by the lowly rodent primary visual cortex", *Learning & Memory*, 21 (10): 527–33.

Glickfeld, L. L. and Olsen, S. R. (2017), "Higher-Order Areas of the Mouse Visual Cortex", *Annual Review of Vision Science*, 3: 251–73.

Glickfeld, L. L., Reid, R. C., and Andermann, M. L. (2014), "A mouse model of higher visual cortical function", *Current opinion in neurobiology*, 24: 28–33.

Hubel, D. H. and Wiesel, T. N. (1962), "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex", *The Journal of physiology*, 160 (1): 106–54.

Hubel, D. H. and Wiesel, T. N. (1965), "Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat", *Journal of neurophysiology*, 28 (2): 229–89.

Hubel, D. H. and Wiesel, T. N. (1968), "Receptive fields and functional architecture of monkey striate cortex", *The Journal of physiology*, 195 (1): 215–43.

Huberman, A. D. and Niell, C. M. (2011), "What can mice tell us about how vision works?", *Trends in neurosciences*, 34 (9): 464–73.

Hung, C. P., Kreiman, G., Poggio, T., and DiCarlo, J. J. (2005), "Fast readout of object identity from macaque inferior temporal cortex", *Science*, 310 (5749): 863–6.

Ince, R. A., Mazzoni, A., Bartels, A., Logothetis, N. K., and Panzeri, S. (2012), "A novel test to determine the significance of neural selectivity to single and multiple potentially correlated stimulus features", *Journal of neuroscience methods*, 210 (1): 49–65.

Intraub, H. (1980), "Presentation rate and the representation of briefly glimpsed pictures in memory.", *Journal of Experimental Psychology: Human Learning and Memory*, 6 (1): 1.

Katzner, S. and Weigelt, S. (2013), "Visual cortical networks: of mice and men", *Current opinion in neurobiology*, 23 (2): 202–6.

Keysers, C., Xiao, D.-K., Földiák, P., and Perrett, D. I. (2001), "The speed of sight", *Journal of cognitive neuroscience*, 13 (1): 90–101.

Kiani, R., Esteky, H., Mirpour, K., and Tanaka, K. (2007), "Object category structure in response patterns of neuronal population in monkey inferior temporal cortex", *Journal of neurophysiology*, 97 (6): 4296–309.

Kobatake, E. and Tanaka, K. (1994), "Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex", *Journal of neurophysiology*, 71 (3): 856–67.

Kriegeskorte, N. and Kievit, R. A. (2013), "Representational geometry: integrating cognition, computation, and the brain", *Trends in cognitive sciences*, 17 (8): 401–12.

Kriegeskorte, N., Mur, M., and Bandettini, P. (2008), "Representational similarity analysis–connecting the branches of systems neuroscience", *Frontiers in systems neuroscience*, 2.

Magri, C., Whittingstall, K., Logothetis, N. K., Panzeri, S., and Singh, V. (2009), "A toolbox for the fast information analysis of multiple-site LFP, EEG and spike train recordings", *BMC neuroscience*, 10 (1): 81.

Mattia, M. and Sanchez-Vives, M. V. (2012), "Exploring the spectrum of dynamical regimes and timescales in spontaneous cortical activity", *Cognitive neurodynamics*, 6 (3): 239–50.

Mead, C. (1989), *Analog VLSI and Neural Systems* (Addison-Wesley VLSI system series; Addison-Wesley), `https://books.google.it/books?id=-j8PAQAAMAAJ`.

Milner, A., Perrett, D., Johnston, R., Benson, P., Jordan, T., Heeley, D., Bettucci, D., Mortara, F., Mutani, R., Terazzi, E., et al. (1991), "Perception and action in 'visual form agnosia'", *Brain*, 114 (1): 405–28.

Mishkin, M., Ungerleider, L. G., and Macko, K. A. (1983), "Object vision and spatial vision: two cortical pathways", *Trends in neurosciences*, 6: 414–7.

Montero, V. M. (1993), "Retinotopy of cortical connections between the striate cortex and extrastriate visual areas in the rat", *Experimental Brain Research*, 94 (1): 1–15.

Munk, H. (1881), *Über die Functionen der Grosshirnrinde: Gesammelte Mittheilungen aus den Jahren 1877-1880* (August Hirschwald).

Nicholls, J., Martin, A., Fuchs, P., Brown, D., Diamond, M., and Weisblat, D. (2012), *From Neuron to Brain* (Sinauer), `https://books.google.it/books?id=eTLzXwAACAAJ`.

Niell, C. M. (2011), "Exploring the next frontier of mouse vision", *Neuron*, 72 (6): 889–92.

Niell, C. M. (2015), "Cell types, circuits, and receptive fields in the mouse visual cortex", *Annual review of neuroscience*, 38: 413–31.

Niell, C. M. and Stryker, M. P. (2008), "Highly selective receptive fields in mouse visual cortex", *Journal of Neuroscience*, 28 (30): 7520–36.

Panzeri, S., Senatore, R., Montemurro, M. A., and Petersen, R. S. (2007), "Correcting for the sampling bias problem in spike train information measures", *Journal of neurophysiology*, 98 (3): 1064–72.

Panzeri, S. and Treves, A. (1996), "Analytical estimates of limited sampling biases in different information measures", *Network: Computation in Neural Systems*, 7 (1): 87–107.

Paxinos, G. and Watson, C. (2007), *The Rat Brain: In Stereotaxic Coordinates* (Elsevier Academic Press), `https://books.google.it/books?id=jSW7MgEACAAJ`.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011), "Scikit-learn: Machine Learning in Python", *Journal of Machine Learning Research*, 12: 2825–30.

Potter, M. C. (1976), "Short-term conceptual memory for pictures.", *Journal of experimental psychology: human learning and memory*, 2 (5): 509.

Prusky, G. T., Harker, K. T., Douglas, R. M., and Whishaw, I. Q. (2002), "Variation in visual acuity within pigmented, and between pigmented and albino rat strains", *Behavioural brain research*, 136 (2): 339–48.

Reinagel, P. (2015), "Using rats for vision research", *Neuroscience*, 296: 75–9.

Rolls, E. T. (2012), "Invariant visual object and face recognition: neural and computational bases, and a model, VisNet", *Frontiers in Computational Neuroscience*, 6.

Rossant, C., Kadir, S. N., Goodman, D. F. M., Schulman, J., Hunter, M. L. D., Saleem, A. B., Grosmark, A., Belluscio, M., Denfield, G. H., Ecker, A. S., Tolias, A. S., Solomon, S., Buzsáki, G., Carandini, M., and Harris, K. D. (2016), "Spike sorting for large, dense electrode arrays", *Nature neuroscience*, 19 (4): 634–41.

Rosselli, F. B., Alemi, A., Ansuini, A., and Zoccolan, D. (2015), "Object similarity affects the perceptual strategy underlying invariant visual object recognition in rats", *Frontiers in neural circuits*, 9.

Rousselet, G. A., Fabre-Thorpe, M., and Thorpe, S. J. (2002), "Parallel processing in high-level categorization of natural images", *Nature neuroscience*, 5 (7): 629–30.

Rousselet, G. A., Thorpe, S. J., and Fabre-Thorpe, M. (2004), "How parallel is visual processing in the ventral pathway?", *Trends in cognitive sciences*, 8 (8): 363–70.

Rubin, G. S. and Turano, K. (1992), "Reading without saccadic eye movements", *Vision research*, 32 (5): 895–902.

Rust, N. C. and 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): 12978–95.

Rust, N. C. and DiCarlo, J. J. (2012), "Balanced increases in selectivity and tolerance produce constant sparseness along the ventral visual stream", *Journal of Neuroscience*, 32 (30): 10170–82.

Sanderson, K., Dreher, B., and Gayer, N. (1991), "Prosencephalic connections of striate and extrastriate areas of rat visual cortex", *Experimental brain research*, 85 (2): 324–34.

Sereno, M. I. and Allman, J. M. (1991), "Cortical visual areas in mammals", *The neural basis of visual function*, 4: 160–72.

Tafazoli, S. (2013), "Behavioral and Neuronal Substrates of Invariant Object Recognition in Rats", PhD thesis (International School for Advanced Studies).

Tafazoli, S., Di Filippo, A., and Zoccolan, D. (2012), "Transformation-tolerant object recognition in rats revealed by visual priming", *Journal of Neuroscience*, 32 (1): 21–34.

Tafazoli, S., Di Filippo, A., and Zoccolan, D. (2014), "Transformation-tolerant perception of visual objects in rats revealed by visual priming".

Tafazoli, S., Safaai, H., De Franceschi, G., Rosselli, F. B., Vanzella, W., Riggi, M., Buffolo, F., Panzeri, S., and Zoccolan, D. (2017), "Emergence of transformation-tolerant representations of visual objects in rat lateral extrastriate cortex", *eLife*, 6: e22794.

Tees, R. C. (1968), "Effect of early restriction on later form discrimination in the rat", *Canadian Journal of Experimental Psychology*, 22: 294.

Tees, R. C. (1972), "Effects of visual restriction in rats on generalization along the dimension of angular orientation.", *Journal of comparative and physiological psychology*, 79 (3): 494.

Tees, R. C., Midgley, G., and Nesbit, J. C. (1982), "Generalization after form discrimination in light-reared and dark-reared rats", *Developmental psychobiology*, 15 (2): 159–70.

Thorpe, S., Fize, D., and Marlot, C. (1996), "Speed of processing in the human visual system", *nature*, 381 (6582): 520.

Vermaercke, B. and Beeck, H. P. O. de (2012), "A multivariate approach reveals the behavioral templates underlying visual discrimination in rats", *Current Biology*, 22 (1): 50–5.

Vermaercke, B., Gerich, F. J., Ytebrouck, E., Arckens, L., Op de Beeck, H. P., and Van den Bergh, G. (2014), "Functional specialization in rat occipital and temporal visual cortex", *Journal of neurophysiology*, 112 (8): 1963–83.

Vinh, N. X., Epps, J., and Bailey, J. (2009), "Information theoretic measures for clusterings comparison: is a correction for chance necessary?", in *Proceedings of the 26th annual international conference on machine learning*, ACM, 1073–80.

Vinh, N. X., Epps, J., and Bailey, J. (2010), "Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance", *Journal of Machine Learning Research*, 11 (Oct): 2837–54.

Vinken, K., Van den Bergh, G., Vermaercke, B., and Op de Beeck, H. P. (2016), "Neural Representations of Natural and Scrambled Movies Progressively Change from Rat Striate to Temporal Cortex", *Cerebral Cortex*, 26 (7): 3310–22.

Vinken, K., Vermaercke, B., and Beeck, H. P. O. de (2014), "Visual categorization of natural movies by rats", *Journal of Neuroscience*, 34 (32): 10645–58.

Vinken, K., Vogels, R., and Beeck, H. O. de (2017), "Recent Visual Experience Shapes Visual Processing in Rats through Stimulus-Specific Adaptation and Response Enhancement", *Current Biology*, 27 (6): 914–9.

Wada, Y. and Yamamoto, T. (2001), "Selective impairment of facial recognition due to a haematoma restricted to the right fusiform and lateral occipital region", *Journal of Neurology, Neurosurgery & Psychiatry*, 71 (2): 254–7.

Wallace, D. J., Greenberg, D. S., Sawinski, J., Rulla, S., Notaro, G., and Kerr, J. N. (2013), "Rats maintain an overhead binocular field at the expense of constant fusion", *Nature*, 498 (7452): 65–9.

Zoccolan, D. (2015), "Invariant visual object recognition and shape processing in rats", *Behavioural brain research*, 285: 10–33.

Zoccolan, D., Oertelt, N., DiCarlo, J. J., and Cox, D. D. (2009), "A rodent model for the study of invariant visual object recognition", *Proceedings of the National Academy of Sciences*, 106 (21): 8748–53.

List of Figures