Chapter 1
General introduction

Although the neurosciences have made a tremendous progress in the last decades several fundamental questions, despite furious debate, are still unresolved. One of these questions is the way information from the outside world is processed in the brain to result in a meaningful behavior. Even if much is known about single neurons and action potentials, there is no definitive answer on the question how the information traveling between neurons is coded. The obvious way information reaches the brain is through sensory systems, of which the visual system is the most important (at least in most mammals). A meaningful behavior can be only be made on the basis of a good perception of the visual scene. But where in the brain is perception located? And how do we recognize an object apart from the background and other objects? If cells have to signal each other whether they belong to the same interesting object there arises a problem. One elegant solution for this *binding problem* is that these cells fire in synchrony, maybe even oscillate with the same frequency. Thus neurons may use a temporal code for perception, or don't they?

Two introductory chapters give an overview of theoretical and experimental work published. Chapter 1 describes the background and evolution of the temporal binding theory and summarizes the main points of critique on this hypothesis. Methodological aspects in relation to experimental support for the binding theory will be discussed in chapter 2. In chapters 3 to 5 research is presented in which several conjectures of this temporal binding hypothesis are refuted. The thesis concludes with a general discussion.

1.1 Organization of the visual system: anatomy and physiology

The importance of the visual system in most animals is reflected in the brain, were a large number of areas, the posterior 50 % or so of the cerebral cortex, are devoted to receive, perceive and process visual information (Zeki 1993). A visual stimulus is received on the retina, sent through the thalamus (lateral geniculate nucleus LGN), to the primary visual cortex, viz. area 17 for anatomists or V1 for physiologists. The knowledge on area 17 or V1 (also known as the striate cortex) and other visual areas is for the largest part derived from anatomical and physiological experiments in cats and primates and will be briefly summarized here.

1.1.1 Intrinsic connections

The visual cortex of the cat grows in the perinatal stages of brain development into anatomically very precise highly specialized brain areas with a very precise topographic order (Luskin and Shatz, 1985). In the classical Hubel and Wiesel model (Hubel and Wiesel, 1962) the cells in the primary visual cortex (area 17) are organized in columns, ocular dominance columns and orientation columns. Both columns combine in hypercolumns. Contemporary research finds cells with shared orientation in pinwheel structures (Bonhoeffer and Grinvald, 1991).

Cells in the visual cortex are highly interconnected: intra-areal with horizontal and vertical connections (Gilbert and Wiesel, 1983). Horizontal connections are also, perhaps more correct, termed tangential connections. In primary visual cortex of most species cells who receive these reciprocal corticocortical connections are organized in patches (Rockland and Lund, 1982). The patches are sized 250-500 µm and separated with a distance comparable with the size of the patches. Cells connect with other patches 1-4 mm away, which is about the same spread as the physiological representation of a single point in the visual field in area 17 (Hubel and Wiesel, 1974). Thus the patchy connections interconnect regions containing cells with overlapping receptive fields (LeVay 1988). The horizontal connections connect cells with similar orientation sensitivity (Gilbert and Wiesel, 1989). In a study using cross-correlation analysis it was found that cell firing was correlated between across several orientation columns (Toyama *et al.*, 1981). Cross correlation analysis is a crucial tool to study the connectivity and/or interactions across cells and will be discussed elaborately in chapter 2. Cross-correlation was found to be strongest between orientation columns of like specificity and possibly eye preference (Ts'o *et al.*, 1986; Hata *et al.*, 1991).

Horizontal connections in cat are developed mainly after birth (Luhmann *et al.*, 1986). Thus, the cortical organization can be artificially changed. Raising kittens in binocular deprivation causes altered horizontal connections (Callaway and Katz, 1991). Strabismic cats develop squint resulting in a segregation of afferents from the two eyes (Singer *et al.*, 1979). In strabismic cat horizontal connections were found to interconnect only ocular dominance columns (Löwel and Singer, 1992), in contrast to normally reared cats where horizontal connections are predominantly related to orientation sensitivity (Gilbert and Wiesel, 1989).

The neuronal population in the cortex exists largely of pyramidal cells (70-80%) and stellate cells. An excitatory (pyramidal) cortical cell receives converging input from approximately 10,000 cells, 85% of which are excitatory synaptic contacts (Braitenberg and Schüz, 1991). More than half of these contacts come from neurons within a 100-200 µm radius of the target cell, reflecting the columnar organization of the cortex. Excitatory cells have only 1-4 diverging synapses with a neighboring cell. Inhibitory cells constitute about 20% of the neural population (Gabbott and Somogyi, 1986). One inhibitory cell contacts 10-40 cells (Kisvarday and Eysel, 1993).

The cortex has a loosely layered structure with six layers. Pyramidal cells are found in all layers, but layer I. Layer IV consists predominantly of densely packed spiny stellate, also called granular, cells. Hence Layers I, II and III are also known as supragranular, layer IV granular and layers V and VI infragranular. Cells are connected vertically (perpendicular) between the different layers of the cortex (Toyama *et al.*, 1981; Grossberg 2001).

1.1.2 Extrinsic connections

Extrinsic connections can be subdivided in feedforward, feedback and lateral connections (Salin and Bullier, 1995). In this hierarchical organization feedforward connections go from lower to higher areas, feedback connections go in the opposite direction with area 17 at the bottom of the hierarchy and lateral connections connect areas on the same level. Experiments in primates show that the visual information travels in a feedforward manner from primary visual cortex (V1) to higher visual areas such as V2, V3, V4, V5, mediotemporal cortex (MT) and inferotemporal cortex (IT). In each area the complexity of information is larger and also the receptive field is larger (Maunsell and Newsome, 1987). More than 30 distinct visual areas have been identified in the primate brain (Felleman and Van Essen, 1991). Visual information is largely split into a dorsal stream which goes to V5 and MT and a ventral stream V2, V3, V4, IT. Traditionally, the dorsal stream is called the ‘where’ path, which is assumed to be largely concerned with motion and location, whereas the ventral stream is called the ‘what’ path, assumed to be concerned with form and color (Ungerleider and Mishkin, 1982). The distinction is later described as ‘guidance of action’ for the dorsal stream and ‘object recognition’ for the ventral stream (Milner and Goodale, 1995).

Feedback connections within the visual system are demonstrated (Rockland and Virga, 1989; Salin *et al.*, 1993). This feedback would provide a top-down control on lower level visual areas as V1 and V2 (Lee 2002; Grunewald *et al.*, 2002).

Whereas feedforward connections are visuotopically organized and terminate in layer IV, but not exclusively, feedback connections are not visuotopically organized and tend to avoid this layer. Lateral connections are characterized by arborization of axonal terminals in all layers. The majority of extrinsic connections is reciprocal.

1.2 Visual perception: localists and connectionists

The crucial question is how the information propagated through action potentials between highly interconnected neurons can lead to a meaningful perception of the surrounding world.

Historically, two views of visual perception, and the brain as a whole, exist: the localist and the connectionist view.

1.2.1 Localists

The localist view stems logically from the anatomy of the visual system. This doctrine is in fact an extension of the hierarchical approach of Hubel and Wiesel that sensory information is first processed in simple cells, which converge to complex cells, which converge to further cells, etcetera (Hubel and Wiesel, 1962). If in each level in the visual system neurons converge to neurons with more sophisticated properties one could deduct a single cell, which represents the entire visual scene. Sherrington coined the term ‘pontifical cell’ for such a cell, but rejected the idea (Sherrington 1941). The concept of a ‘grandmother cell’, invented by Lettvin in 1969 (Barlow 1995) is widely used for the concept of a cell which becomes active when and only when the grandmother, or another specific feature, is present in the visual scene. The main objection against this concept is that an immense number of cells would be required for all possible objects, even more than the estimated 3×10^{10} in the neocortex (Mountcastle 1997). This was most specifically addressed by Harris, who stated that while ‘yellowness’ cells and ‘Volkswagen cells’ may be reasonable, surely specific cells devoted to ‘yellow Volkswagens’ (Weisstein 1973) are not (Harris 1980). Otherwise one could imagine cells for all other combinations (e.g. lime-green Minis) and a combinatorial explosion would occur. Barlow (Barlow 1972) recognized that there must be a more economical way for the percept of an object. To reduce the redundancy of information, i.e. repetition of information in an object, he introduced the concept that the scene is represented by a few ‘cardinal cells’, termed in analogy with pontifical cells. A similar idea is the concept of ‘geons’, forming an alphabet of forms (Biederman 1987). Indication that these highly specialized cells are present are found by several researchers (Parker and Newsome, 1998). Some cells in the inferotemporal cortex are capable of recognizing a distinct face (Desimone *et al.*, 1984). Also the performance of a single neuron in the MT cortex was found to be just as good as the performance of the monkey during a psychophysical discrimination task (Britten *et al.*, 1992).

1.1.2 Connectionists

Meanwhile a connectionist approach emerged. Models of neural networks are studied for decades. McCulloch and Pitts (McCulloch and Pitts, 1943) produced a simple neuron model (integrate- and-fire) and showed that a network of these neurons was capable of computing. Hebb published his influential work on networks, in which synaptic plasticity and cell assemblies were introduced (Hebb 1949). Rosenblatt (Rosenblatt 1961) proposed a feedforward network of perceptrons, consisting of two layers of neurons, unidirectionally bound. The connectionist idea was abandoned with the publication of an influential book on the limitations of perceptrons by Minski and Papert (Minsky and Papert, 1969), and the connectionist view was silent for a decade. At the beginning of the 1980’s a revival of connectionist ideas took place. An influential publication was from Von der Malsburg. He formulated the Correlation Theory in which the spike trains of cells of a dynamical cell assembly are correlated (Von der Malsburg 1981, see next subchapter). This coding is therefore also called *ensemble coding*. Earlier Milner (Milner 1974) had published similar ideas. Critique on localist models focuses on the conclusion that an enormous amount of cells would wait silently until a proper stimulus came along. Coding with these sorts of *smart cells* therefore is referred to as *sparse coding*. In contrast the *coarse coding* in neural networks uses the activity of many interconnected cells to build the visual scene. Coarse coding is radically different from sparse coding in that one single cell does not make a meaningful perception, only a group of cells makes a meaningful perception, and thus the cortex makes use of a *population code*.

Such a distributed representation asks far less cells (Singer 1995; Singer 1999a) The original scene could be deducted from the combined activity of generalized cells. The output of selective cells would be generalized in a next cell, just like simple cells, selective for the same orientation, converge to a complex cell with generalizes for the location (Hubel and Wiesel, 1962). Even Barlow finds this a better alternative to *sparse coding* (Barlow 1993). Although a simple feedforward architecture would be capable of a representation of features explicitly represented in the neurons of a specific population, it would be inflexible and therefore does not explain how a new feature constellation for which no specialized neurons exist would be represented. Therefore a dynamical system of interconnected cells is needed which correlate with each other. The use of a population code thus creates problems of nervous integration: these problems are collectively known as the Binding Problem (Von der Malsburg 1981).

1.3 The Binding Problem

Despite of the attention the Binding Problem receives among neuroscientists, psychologists and many others, it is not a unitary problem, but consists of a class of integration problems on several levels. In his original article Von der Malsburg (Von der Malsburg 1981), where he posed the question on nervous integration, he proposes a Correlation Theory as a solution for binding problems on different levels. On the lowest level, cells belonging to the same assembly are labeled with a temporal code. For that cells have to be capable to distinguish synchrony within a short time frame (milliseconds). If synchronized, they will form a temporary assembly, due to synaptic modulation, termed in analogy with synaptic plasticity, which forms networks in a long time scale (Hebb 1949). This formation of assemblies by temporal binding must be a self-organizing process.

Also, at a higher level, sets of topologically connected cells may be considered as network elements, instead of single cells. These composite elements form a network, where the ensemble of cellular signals form a composite signal. On this level, correlated firing would explain figure-ground separation. Consider a primary visual element as a particular combination of visual qualities from one retinal point: luminance, spectral distribution, direction and speed of movement, etc. All elements with the same qualities would then correlate, and thus form a network. The visual scene will be decomposed into figures. This figure-ground separation is the basis of object recognition.

Because binding is present at different levels it is often not clear what exactly is meant with binding in a particular context. Much of the intricacy in the debate is based on this confusion (see for a set of reviews on the different manifestations of the binding problem the September 1999 issue of Neuron). Therefore I will distinguish several binding problems:

1.3.1 Visual feature linking

At the highest level visual features must be correctly associated. For example, the shape of an object must be associated with the color and location to form a unified representation.