Universität Bremen

Contour Integration: Attentional Effects in a Psychophysics Task and Feature Interactions in a Computational Model

Axel Grzymisch M. Sc.

November 2017

Contour Integration: Attentional Effects in a Psychophysics Task and Feature Interactions in a Computational Model

Faculty of Physics and Electrical Engineering

University of Bremen

A thesis submitted in partial fulfillment for the degree of Doctor rerum naturalium

Dr. rer. nat.

Submitted by

Axel Grzymisch M. Sc.

born in Buenos Aires, Argentina - 9.3.1988

1. Reviewer: Dr. rer. nat. Udo Alexander Ernst

2. Reviewer: Dr. habil. Malte Persike

Submission Date: 13.11.2017

Defence Date: 15.12.2017

Acknowledgement

I would like to thank a number of people for their support, help, fruitful discussions, and mentorship during my time as a Doctoral student. First and foremost, I would like to thank my supervisor, Dr. Udo Ernst. Without your guidance and support I would not have been able to accomplish this task. Thank you for encouraging my research, for always having new and insightful ideas, and for finding interesting projects in which I could pursue my interests. Furthermore, thank you for introducing me and guiding me in my discovery of the field of Computational Neuroscience. Although this field was foreign to me before I joined your group, with your guidance, I have been able to accomplish a monumental task, my Doctoral work.

I would also like to thank Dr. Malte Persike for taking on the role of my second supervisor. Your readiness to share your knowledge, and to provide input into my experimental projects, was invaluable. Furthermore, I appreciate the trust you placed in me and your willingness to become my collaborator. Without you, a large part of my work would not have been possible.

I would also like to thank Dr. David Rotermund for his support as the IT administrator. You saved me in more than one occasion when everything seemed to have gone wrong. Our secretary, Agnes Janßen, has also proven to be invaluable. You helped me deal with all the bureaucratic matters which I came across while at the University of Bremen, and made my life much simpler in many occasions.

My fellow Doctoral students, former and present, thank you for your friendship, for your helpfulness in any matter I came to you with, and for many fruitful discussions (scientific and otherwise) which we shared over the years.

Abstract

In order to achieve object recognition and image segmentation, the visual system is tasked with combining colinear and cocircular edge configurations into coherent percepts. This process is called Contour integration (CI). CI is believed to be a fundamental visual process, psychophysical experiments have shown humans to be remarkably good at integrating contours even when parts of the contours are occluded, or when a contour does not follow a smooth path. Electrophysiological studies have characterized the neural substrates of contour integration. Based on this information, modelling studies have produced algorithms to explain the functioning of putative mechanisms which give rise to CI.

In this thesis, two case studies on contour integration are presented. In the first, psychophysical methods were employed to further characterize humans' ability to detect contours under conditions of ambiguity. In particular, this study introduced a novel method in order to determine whether humans' remarkable efficiency in detecting contours carries over to dynamic scenes. This is an important question given that scenes in nature are highly dynamic, and up to this point, most CI studies have characterized this process in static scenes. It has often been assumed that CI is a stimulus driven process which leads to pop-out percepts. Results from this study challenge these views. They indicate that humans' ability to detect contours deteriorate drastically when shown extended presentations of dynamic stimuli. Furthermore, a set of sub-experiments indicates that top-down processes may play an important role in supporting contour integration under conditions of ambiguity.

In the second case study, a computational model of contour integration was developed in order to account for new psychophysical findings, and further understand the mechanisms underlying these observations. Through a number of psychophysical studies, spatial frequency has been shown to be an important feature on which contours can defined and detected, and which can interact with the process of integrating oriented elements. Thus, a modulation component was added to a structurally simple model of contour integration in order to reproduce these findings. The modulation was based on the assumption that interactions of feature detectors are stronger if their preferred spatial frequencies are similar, rather than dissimilar. Extensive numerical simulations were carried out in order to understand the mechanisms leading to the mentioned psychophysical observations, and to reproduce said psychophysical results.

This thesis presents contributions to the field of contour integration in two areas. In psychophysics, not only do the results from the experiments reported provide support for the emerging idea that CI may be supported by top-down process, but a significant methodological contribution was also made. A new technique to study CI was introduced. This will allow future research to characterize contour integration under new conditions. In the modeling field, a gap was bridged. To the knowledge of the author, the model presented in this thesis is the first to account for the geometrical characteristics of stimuli **and** the spatial frequency component of elements in the stimuli.

Contents

Chapter 1

Introduction

In vision, as Aristotle once said, "the whole is greater than the sum of its parts". In order to perceive the visual world, the visual system must combine a plethora of information. This is no easy task. However, humans, as well as other primates and other animals, are remarkably good in grouping different elements of a visual scene into coherent objects. When looking at a natural image, the ability to distinguish different objects with different properties seems innate, and very rarely does the human visual system err in this task. Humans can easily spot a bird in a tree. If it is a yellow canary, we can effortlessly associate the color yellow as belonging to the bird and not to the tree. We can distinguish the bird from its leafy background, and we can identify it as a bird even when it is partly occluded by branches.

So astonishing is the ability of the human visual system to perform the tasks described above that vision scientist have been concerned with finding the root principles for those processes since the early 20^{th} century. Gestalt psychologists proposed several heuristics which the visual system may use to conduct these tasks (Wertheimer, 1923). Through a number of observations they proposed rules suggesting how humans organize perceptual scenes. However, they did not go as far as to describe the biological substrates for their proposed heuristics due to limitations in experimental techniques. Since then, monumental discoveries regarding the biology of the visual system, and how this may give rise to perception, have been made (Hartline, 1938; Hubel and Wiesel, 1959; Kuffler, 1953; Talbot and Marshall, 1941). It is now commonly accepted that visual perception is rooted in the aggregation of information processed by a large number of different cells. Different cells receive input from different sections of the visual field. The section from which a cell receives its input is called the cell's receptive field (RF). Receptive fields of different cells respond to different properties of the visual input, some respond to oriented edges, some to color, some to texture, etc. In order to perceive a visual scene, this information needs to be combined through a process referred to as Feature Integration. By means of this process, the perception of contours and objects is achieved. Thus, as Aristotle stated, the whole is greater than the sum of its parts.

A central aspect in the field of Feature Integration is Contour Integration (CI). This thesis is focused on the study of contour integration. Contour integration is believed to be a basic visual process involved in the perception of objects' outlines, and the

segregation of figure and ground in visual images. In contour integration, co-aligned or co-circular edge elements are grouped together into the impression of a contour, typically referred to as a contour percept. This process has been extensively studied in vision science. Figure 1.1 shows a prototypical example of a stimulus used in CI studies.

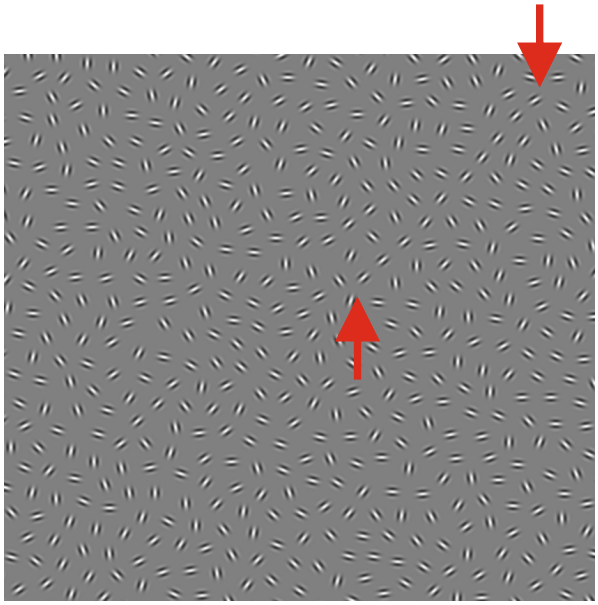

Figure 1.1: Prototypical example of a stimulus used in contour integration studies. A contour is seen on the upper right quadrant, marked by the red arrows.

In contour integration studies, it is often found that edge elements aligned to a global path (and surrounded by randomly oriented edge elements) induce the "pop-out" percept of a contour. That is, observers experience a strong stimulus driven perception of the outline formed by the aligned edge elements, which stands out from the randomly oriented surrounding elements. It is believed that in tasks in which contours "pop-out", no or little, attention is required to perceive the contour. Band-pass elements (often gabor patches), are typically used as edge elements in contour integration studies, as these have properties known to excite the receptive fields of cells in the early visual cortex (V1).

One study of particular relevance to the field of contour integration, and indeed to this thesis, is Field et al. (1993). In their study, after characterizing humans' ability to detect contours under a number of different scenarios (e.g.: deviation of contour elements' orientation to the contour path, varying the separation of elements in a contour, or applying different manipulations to the phase of gabor patches in a stimulus), Field et al. (1993) proposed an association field hypothesis. The association field hypothesis states that feature detectors with RFs in close proximity, and which have similar orientation preferences, have stronger connections than feature detectors with distant RFs, or a dissimilar orientation preference. This hypothesis has been widely accepted and adopted by vision scientists. The association field described by Field et al. (1993) is believed to be rooted in long range horizontal interactions of V1 cells, given that cells in V1 are often connected based on their orientation preferences (Gilbert and Wiesel,

1989), and the proximity of their RFs (Shouval et al., 2000).

Since Field et al. (1993) proposed their association field hypothesis, several studies have been carried out to better characterize humans' ability to perform contour integration, and the mechanisms which might give rise to CI. The effects of closure[1] in contour integration (Kovacs and Julesz, 1993; Mathes and Fahle, 2007), effects of different configuration of contours (i.e.: contour elements aligned to the path angle or perpendicular to the path angle) (Vancleef and Wagemans, 2013), effects of eccentricity (Nugent et al., 2003), temporal requirements for contour integration (Bex et al., 2001), the interplay of alignment and spatial frequency of contour elements (Persike and Meinhardt, 2015a,b; Persike et al., 2009), effects of sharp orientation changes in a contour (Persike and Meinhardt, 2016), and several other effects of stimulus variations in contour integration tasks have been studied. However, most contour integration studies have employed static stimuli to study this process, thus largely ignoring motion, a key component of scenes in nature. Furthermore, although the relationship of orientation cues and similarity cues in CI has been studied, no theoretical framework exists which successfully explains the interactions of these cues.

In this thesis, two fundamental processes in the field of contour integration are studied. Different methods were employed in order to achieve a level of understanding of CI not possible from a single approach. Psychophysical experiments were employed to study contour integration in dynamic scenes. A novel experimental method was developed (Grzymisch et al., 2013) since, to the knowledge of the author, no method existed to study CI under these conditions. Behavioural effects resulting from (possible) interactions of neural processes occurring in high order visual areas (e.g.: V4), and non-visual areas (e.g.: prefrontal cortex), were quantified. The aim of the work carried out with this newly developed stimuli was twofold. First, to address the effects of dynamics in the process of contour integration since CI has almost exclusively been studied with static stimuli. However, scenes in nature are dynamic and ever-changing. Thus, contour integration in dynamic scenes is an important topic which needs to be addressed. Second, to provide a greater understanding of how processes originating in areas other than the early visual cortex (i.e.: V1) may affect the detection of contours, and thus, the perception of simple visual stimuli.

As illustrated with the example of a bird sitting on a branch, in vision, there are several cues involved in the the task of object recognition. In the second part of this thesis, empirical and theoretical work was combined in order to create a mathematically well defined framework capable of explaining the interactions of alignment and similarity cues (in terms of element's spatial frequency) in CI. A large collection of empirical results on the interaction of spatial frequency and alignment cues in contour integration already exists. However, prior to its introduction in (Grzymisch et al., 2016) no theoretical framework capable of explaining the mechanisms which give rise to these observations existed in the literature. With the aim of creating this framework, a computational model was developed, using the modelling of neural networks. The model was based on known mechanisms of the visual cortex. Thus, by processing images with similar mechanisms as those employed by the visual system, inferences could

[1] A closed contour being one that encloses an area in space as opposed to an open one that does not.

be made on how these different mechanisms may interact, how they may be tuned, and in turn, how the visual system may process these same images.

In essence, this thesis revolves around two important questions in the field of contour integration:

1. What are the effects of a dynamic history in a contour integration task, and how do top down processes affect contour integration in cases of high ambiguity.

2. Can a contour integration model which accounts for feature similarities, in terms of orientation and spatial frequency, be built to reproduce psychophysical results. In particular, can the model be conceived around the principle that interactions of feature detectors are stronger if their preferred spatial frequencies are similar, rather than dissimilar, and if their RFs are visuotopically close, rather than far.

In this thesis, these two questions will be answered with the aim of bridging a gap in the field of contour integration. A novel experimental method addressing the role of motion, a key component in natural scenes, will be discussed, and a unifying theory capable of explaining a large pool of empirical observations will be introduced.

The content of this dissertation is organized in the following manner. First, an introduction into basic concepts is provided. In this first section, the anatomy and functional properties of the visual system are described, followed by an overview of Gestalt principles and their neural correlates.

Second, a novel experimental paradigm designed to study contour integration in dynamic scenes is introduced. Relevant literature in CI is discussed, and the motivation for a set of experiments is provided. A report of two experiments follows. The results are discussed and a conclusion detailing the implications of the findings, alongside with potential neural mechanisms leading to said findings is provided. Note that this chapter closely follows a journal article published on this topic (Grzymisch et al., 2017a), and that data related to these experiments was presented at conferences prior to the article's publication (Grzymisch et al., 2013, 2015).

Third, a model of feature integration which accounts for orientation and spatial frequency similarities is reported. Earlier versions of this model have been presented at several conferences (Grzymisch et al., 2016, 2017b,c), and a journal publication of this model is currently in preparation. The psychophysical results motivating the model are summarized, and a literature review discussing existing contour integration models ensues. A detailed description of the model, and the methods used to evaluate the model's results is provided, followed by a report on numerical simulation results. The chapter ends with a conclusion outlining the accomplishments of this model, and with a discussion of future research directions.

The dissertation concludes with a general discussion elucidating the advances brought by the work presented in this thesis, and possible research directions involving both the psychophysical experiments and model presented in this thesis.

Chapter 2

Basic Concepts

2.1 Introduction

Sight, the ability to see, is arguably the most important of the five senses (sight, taste, touch, smell, and hearing) for primates. A testament to the importance of vision in primates is the devotion of one of the four major lobes of the cerebral cortex, the occipital lobe, to visual processing (Brown and Schafer, 1888). Because we rely so heavily on vision we often assume this sense to be ubiquitous throughout the animal kingdom. However, there are several species which have very different visual systems than ours, and there are other species which do not possess eyes and have so far managed to survive (e.g.: the Texas Blind Salamander (Eurycea rathbuni), the Blind Tetra (Stygichthys typhlops), etc.). We find the lack of vision in other species to be such a defining feature that we often assign the word "blind" into their colloquial names. However, we must remember that we, as all other species, have evolved under a particular environment and our senses have adapted to our habitat.

In this chapter, a number of basic principles required to understand the remaining of this thesis will be discussed. The reader will first be provided with a basic overview of the anatomy and functional properties of the primary visual cortex (V1). Neurons of particular relevance will be described in detail. It is important to note that this is a basic overview. Thus, it is assumed that the reader already possesses enough biology knowledge in order to understand this discussion. If the reader lacks this knowledge, it is recommended that he consults a neuroscience textbook (e.g.: Kandel et al. (1991)). Following the aforementioned synopses, the importance of vision in terms of our everyday experience of the world is discussed. Armed with this knowledge, the reader is presented with a discussion of different functional aspects of other cortical processes (e.g.: attentional modulation effects), and a description of their implications in visual processing is provided.

Finally, the topic around which this thesis is centered, Feature Integration, is introduced and related to the concepts previously discussed in this chapter.

2.2 Vision

Vision differs from sight in that sight is the ability to see, whereas vision is the ability to organize, interpret, and understand what one sees. As such, this thesis is concerned with vision rather than sight.

The process of vision starts in the eyes, when photons strike photoreceptors located in the back of the ocular globes. There are two types of photoreceptors: rods and cones. These take their names from their shapes, but they also differ in their anatomy and functionality. Their main differences are summarized in table 2.1.

Rods	**Cones**
High sensitivity: responsible for vision under dim light conditions (more photopigments than cones)	Low sensitivity: responsible for vision under normal (day) light conditions (less photopigments than rods)
Saturates in daylight	Graded response under different light intensities (saturates only in intense light)
Low temporal resolution (∼12 Hz)	High temporal resolution (∼55 Hz)
More sensitive to scattered light because of its shape	Most sensitive to directed axial rays because of its shape
Low visual acuity: high conversion of individual receptors into bipolar cells, not present in central fovea	High acuity: less conversion of individual receptors into bipolar cells, concentrated in the fovea
Achromatic: only one type of rod pigment	Chromatic: three types of cones, each with a different pigment that is more or less sensitive to a given part of the light spectrum

Table 2.1: Property differences between rods and cones, and between their neural systems (Kandel et al., 1991).

Note that rods and cones project to bipolar cells, these in turn project to another family of cells called Ganglion cells. Ganglion cells are of particular importance because they are not only the "output" neurons of the retina, but also because they provide a window to study how the retina responds to different patterns of light (Curcio and Allen, 1990; Lee et al., 1990).

Ganglion cells' axons exit the eye through the optic never and carry all visual information to higher visual centers in the brain (Dowling, 2009). Each ganglion cell in the retina has a specific receptive field. That is, each ganglion cell responds (by either increasing or decreasing its firing rate) to light that strikes the retina in a specific area (Wiesel, 1960). The three main features of ganglion cells' receptive fields are that: **a.** they have a (roughly) circular receptive field Kuffler (1953); **b.** they often[1] have a

[1]Some ganglion cells also respond to changes in the overall luminance of the visual field and are

circular center (the receptive field center) and an antagonistic ring shaped surround. **c.** The receptive fields are either **on-center** or **off-center**. The former is excited when presented with a luminous stimulus in the center, and no luminous stimulus in the surround; and the latter is excited when presented with a luminous stimulus in the surround, and no luminous stimulus in the center (Kuffler, 1953); and **d.** they process information in two parallel pathways. That is, information from distinct photoreceptors is not directed to either on-/off-center ganglion cells, but rather to both of these types of cells (Kandel et al., 1991).

After Ganglion cells' axons exit the eye through the optic disc they bundle together to form the optic nerve (Kandel et al., 1991). The optic nerves from the two eyes first meet at the optic chiasm. At the optic chiasm the fibers from each hemifield[2] are sorted into the left and right optic tracts. The optic tracts project to the lateral geniculate nucleus (LGN) in the thalamus.

The LGN is the is the first stage in the brain[3] for visual information processing. The lateral geniculate nucleus is divided into the right and left LGN. The left optic tract projects to the left LGN, and the right optic tract projects to the right LGN. There is a visotopic representation in the lateral geniculate nucleus, since ganglion cells from the retina project in an orderly manner to the LGN. Although a visotopic representation exists in the lateral geniculate nucleus, this does not mean that there is a proportional correspondence between the surface of the retina and the LGN. Rather, in the lateral geniculate nucleus, areas with the highest acuity of vision (thus of density of photoreceptors) are allocated a bigger area. Approximately half of the neural mass in the LGN (and in the primary visual cortex) represent the fovea and regions around the fovea[4] (Kandel et al., 1991). See figure 2.1 panel **A** for an illustration of visual information flow from the retina to the primary visual cortex, and panel **B** for an anatomical view of the the LGN layers.

Receptive fields in the LGN are very similar to those of the retina. They tend to have a circular shape, they have an on or off center, and an antagonistic surround (Hubel and Wiesel, 1959). This similarity arises since each cell in the LGN receives input originating from a few ganglion cells (Cleland and Levick, 1974).

The exact function of the LGN is still unclear. About 10-20% of the presynaptic connections onto the geniculate relay cells are from the retina. This means that most connections in the LGN originate from elsewhere, in many cases from the cortex. Thus, indicating that there are feedback loops already in the LGN, and that these feedback loops may control the flow of information to the cortex (Kandel et al., 1991).

---

important in controlling pupillary reflexes (Kandel et al., 1991).

[2]The retina of each eye is divided into the temporal hemiretina and the nasal hemiretina. These are two sections in the retina which are divided at the optic disc. The right visual hemifield is composed of light which strikes the temporal hemifield of the left eye and the nasal hemifield of the right eye, and vice versa for the left visual hemifield (Purves, 2004).

[3]"Brain", referring to brain matter located in the cerebrum skull cavity.

[4]The fovea is an area at the back of the eye where light is mainly focused by the cornea and the lens. A high concentration of cones are found in this region, giving the fovea a higher visual acuity than other areas in the retina.

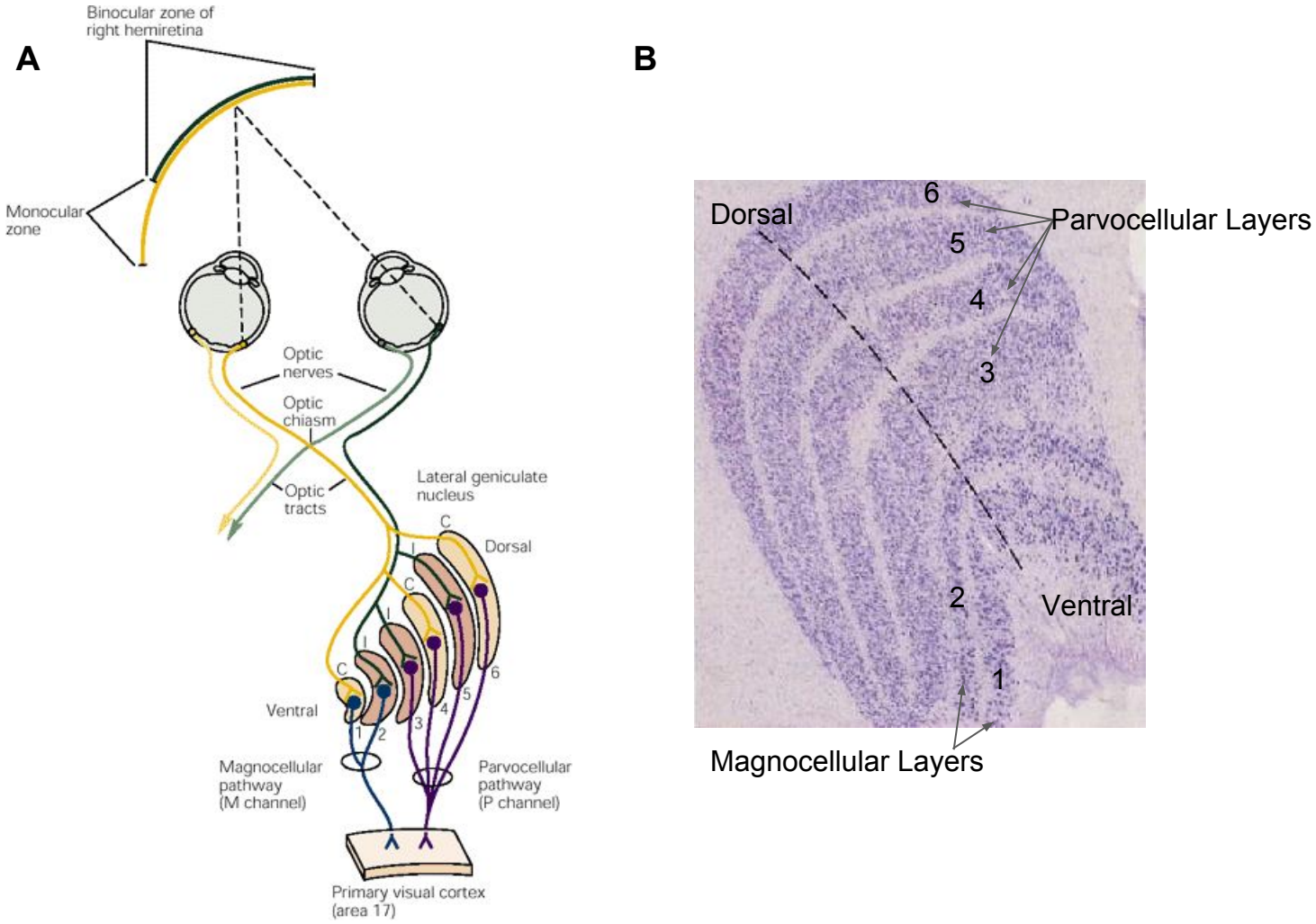

Figure 2.1: **A.** Diagram of the Lateral geniculate Nucleus (LGN). Depiction of inputs from the right hemiretian projecting to different layers in the LGN (left not shown). Magnocellular layers: 1 and 2; Parvocellular layers: 3 to 6. Output of all LGN layers project to the primary visual cortex. Letters C and I for the different layers of the LGN indicate which layers receive contralateral input or ipsilateral input, respectively. Image obtained from (Kandel et al., 1991). **B.** Nissl stain of an LGN section cut parallel to the face of a macaque monkey. Layers 1 through 6 are labeled and the Magnocellular and Parvocellular layers are identified. The Dorsal and Ventral positions of the LGN are identified. Image obtained from (Hubel, 1995).

2.2.1 The Visual Cortex

The visual cortex is the area in the brain in which most of the processing of visual information occurs. It is located in the occipital lobe (see figure 2.2), and it is divided between both brain hemispheres (the corpus callosum connects it). The hemispheres receive input from the ipsilateral LGN, meaning that they process information from their contralateral visual field (i.e.: the left hemisphere receives and processes input from the right visual field, and vice versa).

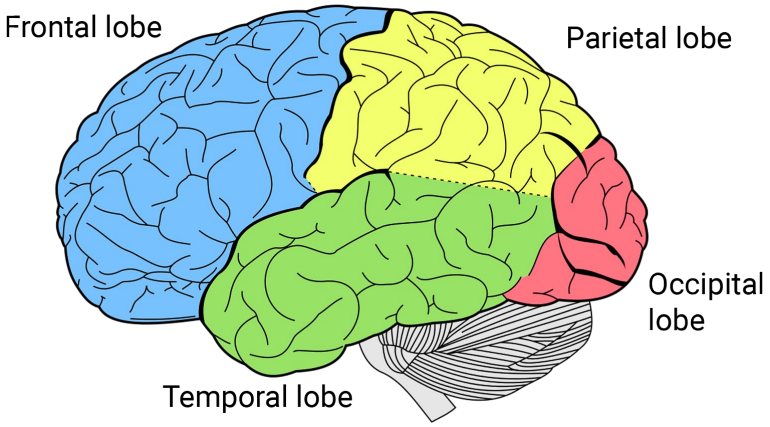

Figure 2.2: The four major lobes of the brain are shown in different colors. The frontal lobe is responsible for higher executive functions; the parietal lobe is responsible for integrating a number of distinct types of sensory information (e.g.: touch, temperature, pressure, and pain); the temporal lobe is responsible for the processing of sensory information, in particular for the processing of auditory information. However, certain aspects of visual processing (e.g.: faces) have also been observed to take place in the temporal lobe; the occipital lobe is the major visual processing center in the brain. Note that the description of specialization of the different brain areas is a rough guideline rather than a set rule. Image obtained from (Institute, 2016).

The visual cortex is divided into different areas where information is processed in a mainly (except for feedback loops) hierarchical manner (Clarke and Miklossy, 1990). As a rule of thumb, one can think of "lower" areas in the visual cortex processing information of a lower complexity, and vice versa. As the processing progresses through the different layers of the visual cortex, the type of information being processed increase in complexity. As in the LGN, there is also a retinotopic/visotopic map in the primary visual cortex (V1). That is, points that are adjacent in the visual field also correspond to adjacent areas on the cortex (Adams and Horton, 2003). A further subdivision of the visual cortex can be made on the basis of striate cortex (V1) and the extrastriate areas, consisting of V2 to the middle temporal area (MT), sometimes also referred to as V5 (particularly in non human primates). The primary visual cortex is of most relevance for this thesis, thus, a detail discussion of the cells which make up V1, its architecture, and functional properties ensues.

The primary visual cortex is where the structure of an image begins to form. V1 receives most of its input from the LGN, and sends most of its output to subsequent cortical visual areas (there are feedback connections from V1 to subcortical areas, however, focus will be placed on V1 and higher visual areas in this discussion) (Maunsell and Newsome, 1987). There are about 40 neurons in V1 for every LGN neuron (Wandell, 1995), meaning that a significant amount of convergence already takes place at this stage of the visual cortex. The two most prominent types of neurons in V1 are simple and complex cells (Hubel and Wiesel, 1959, 1962). These types of cells are of such importance in the visual system that they merit a discussion section of their own.

2.2.1.1 Simple and Complex Cells

The names of simple and complex cells arise due to the complexity of their receptive fields. Simple cells have receptive fields that are mainly activated by oriented edges or line segments. They are tuned to particular orientations, meaning that a particular simple cell responds the most to an edge, or oriented bar, at a particular angle. As this angle deviates from a cell's preferred orientation, its firing rate decreases (Hubel and Wiesel, 1959). This gives rise to a tuning curve (see figure 2.3). Note that simple cells' tuning curves peak at a certain angle, and (typically) decay symmetrically with deviations from said angle.

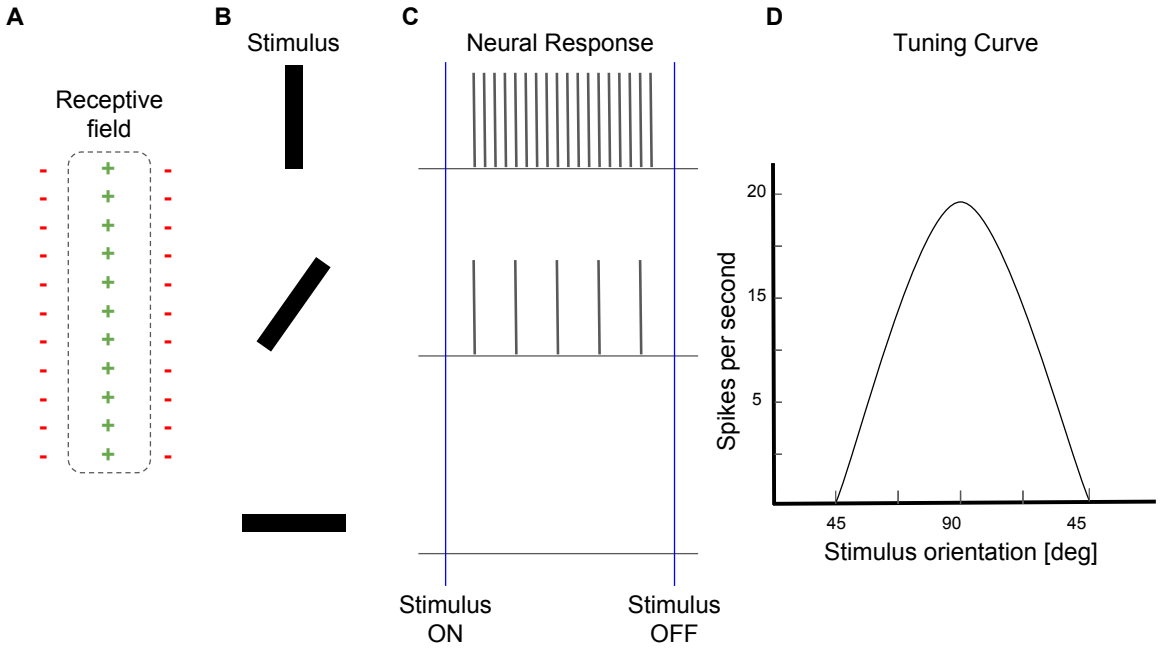

Figure 2.3: **A**: Sketch of the Receptive field of a neuron (red minuses indicate inhibitory area; green pluses indicate excitatory area). **B**: Stimuli presented to the neuron (bars of different orientations). **C**: Recorded neural responses of the stimulus when presented with the stimuli in **B**. **D**: Tuning curve of the neuron. By presenting the neuron with a number of different orientations and recording its response to each of the different orientations, a tuning curve can be derived to show for which orientations the neuron presents the highest response. In this stipulated case, the neuron would respond maximally to a bar oriented at 90°. Note that this is not a real tuning curve of a real neuron but rather a sketch to show how a tuning curve could be obtained. Image adapted from (Goldstein, 2009).

As shown in figure 2.4, simple cells have inhibitory and excitatory regions. The RFs of these cells are elongated. Hypothetically, this elongated shape arises due to the summation properties of a simple cell. That is, a simple cell receives input from several cells in the LGN, and since cells in the LGN have circular receptive fields, they can be arranged in the manner shown in figure 2.4 to form an elongated receptive field. The inhibitory and excitatory region of a simple cell act in an antagonistic manner, meaning

that when there is diffuse light (which falls on both the excitatory and inhibitory regions of a cell's RF), rather than a bar of light simulating only the excitatory region, the cell shows very little activity (Hubel and Wiesel, 1959).

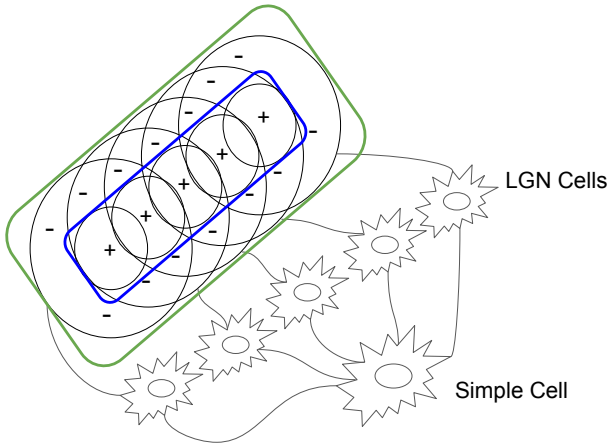

Figure 2.4: Sketch of a simple cell's receptive field. The full receptive field of a simple cell is encapsulated in the green rectangle. The receptive field is composed of an inhibitory section (area outside the blue rectangle and marked with minus signs) and an excitatory section (area inside the blue rectangle and marked with plus signs). The receptive field of a simple cell has an elongated shape. This shape arises due to the summation properties of simple cells. The receptive fields of adjacent LGN cells overlap with each other in the manner depicted, to form an elongated receptive field when their input is aggregated BY one simple cell.

Complex cells primarily receive input from simple cells (Hubel and Wiesel, 1962), and to a much smaller extent from cells in the LGN (Palmer, 1999). They have larger receptive fields than simple cells, and like simple cells, complex cells also (weakly) respond to oriented bars or edges. However, the response of complex cells is not as highly dependent on the position at which an oriented bar appears on its receptive field, as is the case with simple cells (Hubel and Wiesel, 1962). The reason for this is that complex cells receive input from several simple cells (thus they have several adjacent excitatory regions in their receptive field, and their RFs are larger). As a result of having adjacent excitatory regions, complex cells also respond particularly well to movement across their receptive fields. That is, if an oriented bar moves across the visual field without changing its orientation, this movement will elicit a strong response in a complex cell. The sketch in figure 2.5 shows that this particular complex cell receives input from several simple cells which have a 90° preference to an oriented bar of light. If the bar of light shown on the left moves from left to right through the receptive field of the complex cell, then it would activate all four simple cells at subsequent time intervals. These four simple cells would relay their input to the complex cell (each with a short delay when compared to the previous one), leading to a stronger response than if the bar of light would be static in the receptive field of the complex cell. The stronger response would be due to the integration time of the

complex cell being greater than the delay between the time in which each of the simple cells relays its signal to the complex cell. Thus, the response of the different simple cells could be integrated by the complex cell.

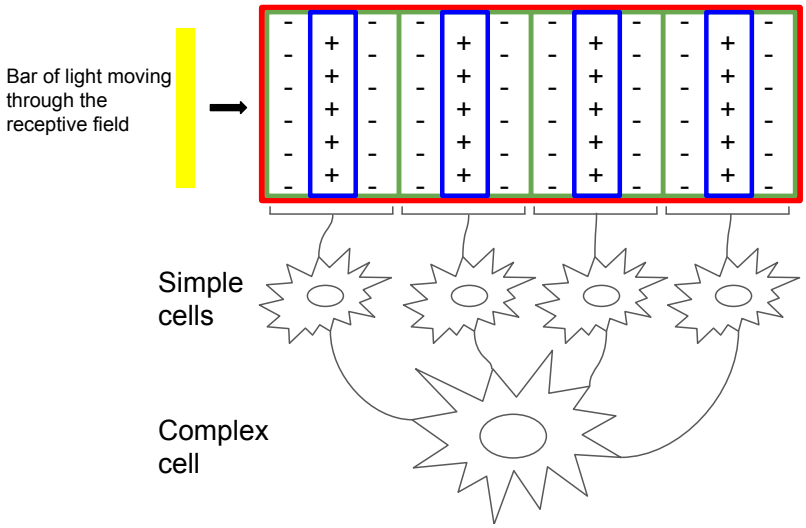

Figure 2.5: Sketch of a complex cell's receptive field. The full receptive field of the complex cell is encapsulated in the red rectangle. The receptive field is composed of adjacent receptive fields of simple cells. This leads to less specificity in the position where a bar of light must lie on the receptive field to evoke a response in a complex cell than when compared to a simple cell (since there are several excitatory regions). It also leads to complex cells being strongly activated by moving edges or bars of light. When the bar of light on the left moves through the receptive field of the complex cell it will first pass through the individual receptive fields of the simple cells which make up the RF of the complex cell, and thus activate them. In turn, the simple cells will relay their input to the complex cell, thus leading to a stronger activity patter than if the bar of light was static, since (in this case) the signals from four simple cells would be integrated by the complex cell.

Simple and complex cells in V1 are not only orientation and direction selective, but they can also have selectivity for spatial frequency. That is, some cells are most excited by gratins of a certain spatial frequencies (Andrews and Pollen, 1979; De Valois et al., 1982). The selectivity for different spatial frequencies arises by a cell having different numbers of inhibitory and excitatory regions interlayed between each other (Carandini, 2012). As with orientation and direction of movement, cells which are selective to spatial frequency also have tuning curves for their spatial frequency preference (Hamilton et al., 1989). Furthermore, the spatial frequency preference of neurons is also correlated to other properties of a cell, such as the size of the receptive field. Neurons with a large receptive field usually have a low spatial frequency preference (Nienborg et al., 2004).

The key points to note are that V1 is composed of a large number of simple and complex cells which have a spectrum of orientation, spatial frequency, and direction preferences. All possible orientations, directions of movement, and visible spatial fre-

quencies are covered by cells in V1 (Hubel and Wiesel, 1974). Given that all possible orientations are represented in V1 by distinct cells, the outlines of a visual image can be decomposed into short line segments of various orientations. Thus, the structure of visual images can be coarsely decoded from activity in V1 (Movshon and Simoncelli, 2015). As noted previously, in contour integration studies edge elements are often realized as Gabor patches. These type of gratings are used as their properties match those of cells in V1. Thus, by generating images with Gabor patches, stimuli can be design so that they are suitable to be processed by V1.

2.2.2 Architecture and Functional Properties of the Primary Visual Cortex

V1, the primary visual cortex, is divided into six layers (layers 1 through 6), and each of these is functionally different (Hubel and Wiesel, 1962). Layer 4 is where most of the LGN neurons synapse, and this layer can be further divided into four sub layers (layers 4A, 4B, 4Cα, and 4Cβ). Layer 4Cα mainly receives input from the LGN's magnocellular pathway, and 4Cβ mainly from the LGN's parvocellular pathway (McLaughlin et al., 2005).

V1 has a very well defined retinotopic map, and as previously mentioned, it has cells which are tuned to respond to different orientations, direction of movement, and even spatial frequency (Hubel and Wiesel, 1959, 1962). In V1, as well as in other cortical layers, there are "orientation columns". These are organized regions of neurons which are excited by oriented line stimuli. The orientation columns are physically arranged in a geometry in which the different slabs making up a column represent slightly different orientations in a circular continuum (Hubel and Wiesel, 1974). This architecture may intensify orientation selectivity, as neighboring cells (with a different orientation selectivity) may provide lateral inhibition.

In all layers of the visual cortex an organization pattern has been observed, where cells with similar tuning curves are strongly connected to one another. Also, cells with very similar tuning preferences tend to have receptive fields which have a great degree of overlap. This arrangement allows for all orientations and motion directions to be represented in the visual cortex in a coherent manner.

Field et al. (1993) proposed their association field hypothesis due to the architecture and functional properties of V1. As previously mentioned, the association field hypothesis states that feature detectors with receptive fields in close proximity, which have similar orientation preferences, have stronger connections than feature detectors with distant RFs (or a dissimilar orientation preference). This hypothesis was used to explain seminal results in the field of contour integration (Field et al., 1993), as such, it has guided much of contour integration literature - including work to be presented in this thesis.

For the readers of this thesis, it is important to note that: **a.** there are cells in V1 which are known to be excited by oriented stimuli; **b.** neurons with a similar tuning curves share strong connections; **c.** the tuning curves of simple cells peak at a certain angle and tend to decay symmetrically with deviations from said angle; **d.** the receptive fields of neurons with similar tuning curves tend to be in close visuotopic

proximity; and **e.** the receptive field size of V1 neurons tends to be correlated to a neuron's spatial frequency preference.

2.2.3 Attentional Modulation

The retina receives an incredibly large amount of input. In order to make sense of our visual environment we must select a subset of said input, as otherwise we would likely be overwhelmed by the large amount of information. There are different ways in which this subset can be selected. Some are stimulus driven, meaning that the stimulus itself draws our focus towards it; and others are goal oriented, meaning that we select this subset based on the task at hand. Attentional modulation driven by goal oriented behaviours is referred to as "top-down attention", as it is thought to originate in high order cortical areas. Top-down attention can be controlled voluntarily, thus it is very well suited to study the effects of attention on vision. As such, many experiments which focus on the effects of attention involve top-down attention. Typically, in these studies neural responses to an attended visual stimulus are compared with the neural responses to an unattended visual stimulus. Generally, what is found is that a stimulus which is attended elicits an enhanced neural response when compared to the unattended stimulus. Several studies (Chelazzi et al., 1993; Moran and Desimone, 1985; Reynolds et al., 1999) have shown that under unattended conditions, the neural response of a single receptive field to two distinct stimuli (presented simultaneously) is similar to the average response observed when the two stimuli are presented individually. However, under conditions in which one of the two simultaneously presented stimuli in the receptive field is attended, the neural response observed is similar to that seen when an individual stimulus is presented (as if the unattended stimulus were not present). This has been dubbed bias competition by Desimone and Duncan (1995). In their paper, Desimone and Duncan (1995) proposed that when two (or more) visual stimuli are presented simultaneously, the neural responses observed are determined by the responses of the two stimuli acting in a "competitive" manner for neural resources. The winner of this competition will be allocated more neural resources to its processing than the other candidate(s). Thus, the competitive interactions can be biased to preferentially select one of the visual stimuli in a scene.

Attentional modulation has been shown to have several other effects in visual processing. When there is only one item in a receptive field, neural responses are higher if attention is allocated to the item inside the receptive field, rather than elsewhere (Treue and Trujillo, 1999). Furthermore, the tuning curves of neurons in a receptive field can also be affected by attentional modulation. McAdams and Maunsell (1999a,b) found that tuning curves of orientation selective neurons in V4 are enhanced (i.e.: their standard deviation decreases) when rhesus monkeys pay attention to an oriented stimulus in the receptive field being recorded, rather than to a stimulus outside of the receptive field. That is, the ability of the recorded neurons to discriminate between similar orientations was enhanced under attentional conditions.

Improvements in performance similar to the ones described above have also been observed when brain areas associated with attention are electrically stimulated. When the frontal eye fields (FEF) of monkeys performing a task in which they were required

to detect the dimming of a visual target were stimulated with subthreshold currents, a behavioural improvement was seen (Awh et al., 2006). In a separate study, Awh et al. (2006) also found that when stimulating superior colliculus neurons, monkeys are better able to discriminate direction changes of a visual stimulus.

Most attentional modulation effects are seen in high order visual areas, such as V2, and V4. However, some (e.g.: Luck et al. (1997)) have argued that they are also present in V1, but are harder to evaluate. Feedback connections from higher visual areas have been shown to have effects on V1 neurons (Hupé et al., 1998). Thus, indirect attentional modulation effects can also be seen in V1. Haenny et al. (1988) presented monkeys with grated patterns and the responses produced by the last pattern (the only reward contingent stimulus in the sequence), were compared with responses obtained from the same stimulus earlier in the sequence. They found that attentional modulation effects led to a response increment of 20% in V1 neurons, whereas the same effect led to a response increment of 72% of V4 cells. Because of the difference in response increments between these two areas they hypothesized that the increments seen in V1 were driven by higher order feedback systems. In contour integration studies, temporal observations of enhanced responses in different areas of the visual system also suggest that V1 is not directly influenced by contextual modulation effects, but rather by feedback connections from higher order visual areas, which are directly affected by attentional modulation (Gilad et al., 2013).

In conclusion, attentional modulation effects exist throughout the visual cortex and they help modulate our perception of visual stimuli. The strongest effects of attentional modulation in the visual cortex are seen in high order visual areas. However, there is strong evidence suggesting that due to feedback loops, the processing of visual information in low visual areas (such as V1) is also affected by attentional modulation, whether directly or indirectly.

2.3 Feature Integration

The content of this thesis revolves around the topic of Feature Integration. Thus, in order to understand the remaining chapters of this thesis readers need to be armed with the understanding of this concept. In the following section a basic introduction into the concept of feature integration and the history of the field will be provided.

2.3.1 Gestalt Principles and Their Underlying Neural Correlates

When viewing a visual scene, we receive a large amount of input. Our experience of the visual world suggests that we can effortlessly see complex objects which are composed of several individual components. For example, as I am writing this thesis, I can look down and see a computer keyboard. If I want to describe in greater detail what I actually see, I can describe this visual image by describing the individual keys. If I wanted to break down this image even further, I could describe the edges of each of the keys which composes the keyboard, and I could continue down this path until

the smallest component which I can consciously perceive. In this manner one can appreciate that what we assume to be an object in our visual world is composed of a large number of features which give rise to our perception of said object. We group, or "integrate", these feature in order to create our visual experience of the world, and indeed, also to make sense of it.

Several heuristics describing how we integrate features to arrive a the percept of an object/figure have been proposed by Gestalt psychologists. The main heuristics include the Law of Good Continuation, which states that elements which follow a smooth global trajectory are perceptually grouped; the Law of Proximity, which states that elements tends to be perceived as belonging to a group if they are in physical proximity; the Law of Similarity, which states that elements tend to be grouped together if they are similar to one another (their similarity can be conceived in may different ways, e.g.: by color, shape, spatial frequency, etc.); and the Law of Common Fate, which states that elements that move together tend to be perceived as belonging to the same group (Coren and Girgus, 1980; Wertheimer, 1923). The former three are of most relevance for readers of this thesis, as theories of contour integration are often (partly) based on these heuristics (Wagemans et al., 2012a). As such, these three heuristics will be discussed in detail.

Although the grouping heuristics proposed by Gestalt psychologists present simple concepts, once one delves into them, important questions arise. For the Law of Good Continuation, which has mainly been studied with artificial stimuli (e.g.: images generated with Gabors, illusory contours such as the Kanizsa triangles, etc.) one such question is how ecologically valid this principle might be. In the last two decades, evidence of collinearity and cocircularity in the statistics of natural images has been found (Krüger, 1998; Sigman et al., 2001). Thus, the method to study contour integration developed by Field et al. (1993), and those employed by Gestalt psychologists to study good continuation, are apt to extrapolate the validity of their results to natural images.

For the Law of Proximity the question of ecological validity is also relevant, however, a question regarding the relationship between proximity and grouping strength is also interesting. Several studies have tried to answer the question of how does grouping strength vary as a function of the separation of elements. The most commonly accepted proposal is that this relationship is described by a power law (Kubovy et al., 1998; Oyama, 1961). Kubovy et al. (1998) also found that this rule is (approximately) scale invariant. That is, scaling all distances by the same factor did not affect their results. As discussed by Wagemans et al. (2012a), this result adds validity to the power law proposal, as this is the only perfectly scale-invariant relationship which has been proposed to describe the relationship between grouping strength and distance. Ecological validity for this heuristic has also been found. Sigman et al. (2001) have found that a power law is found in the spatial correlation in the response of collinearly-oriented filters to natural images.

As for the Law of Similarity, what has mainly been explored is how this heuristic interacts with others. For contour integration, it has generally been found that performance decreases as elements differ in one (or more) of their physical characteristics (e.g.: the phase or spatial frequency of Gabor elements) (Wagemans et al., 2012a). Ecological validity has also been found for the Law of Similarity. Elder and Goldberg

(2002) studied the difference in brightness between edges in natural images. Their findings revealed that similarity, on the basis of brightness of an element, is indeed an important grouping cue.

There are other grouping heuristics which have been proposed, however, these work at higher perceptual level than the heuristics previously described. The Law of Past experience and the Law of good Gestalt are two examples of these. Although these grouping principles are important, only feature integration which is driven by the properties of a stimulus will be discussed in this section, as this is most relevant to contour integration. Other characteristics of stimuli which aid in contour integration have been observed, but these tend not to be as general as the three heuristics previously mentioned. Closure and symmetry are two such characteristics which shall also be described.

Close contours leads to a different percept than an open ones. Close contours, Koffka (1935) argued, generate the percept of a figure (or object) which can be discerned from the ground. Thus, by having a close contour a special percept arises. Open contours, on the other hand, appear as lines rather than as areas in space. Modern contour integration literature has found that close contours, with the same (or similar) statistics than open contours, lead to better contour detection (Mathes and Fahle, 2007; Persike et al., 2009; Sakamoto et al., 2008). However, some (e.g.: Tversky et al. (2004)) have also argued that the better contour detection performance given by close contours may be due to a "better continuation" of the contour. As such, this still remains a debated issue.

Symmetry was identified by gestalt psychologists as one of the characteristics of a "good shape" gestalt. It was not identified as one of their main heuristics, and can possibly be classified as a sub-category of other perceptual characteristics known to induce grouping (e.g.: convexity). Thus, although symmetry might not seem that important, it has gain a great degree of relevance with the advent of computer vision. Symmetry has been used in contour integration algorithms (e.g.: Stahl and Wang (2008)), and it is also an important factor in the contour integration algorithm presented in Chapter 4 of this thesis.

As it can be appreciated from the discussion above, feature integration can occur in many different ways. As such, it is important to have an understanding of how the different grouping principles suggested by Gestalt psychologists could be supported by biology.

The Law of Good Continuation may find neural correlates in long range connections between neurons with orientation tuning preferences of similar orientations, as suggested by (Hess and Field, 1999). The Law of Proximity may be based on neurons with distinct receptive field sizes. As mentioned in section 2.2.2, neurons that are higher up in the visual cortex hierarchy tend to have receptive fields which are larger than those lower in the hierarchy. This means that if a a subject is presented with a set of three edge elements with some separation between them, but with the same orientation, V1 neurons may be able to distinguish between these three lines. However, a neuron with a larger receptive field, for instance in V4, may react to the line as a whole object (thus grouping it), as it can encompass all three elements within its receptive field (Todorovic, 2011). The grouping principle of similarity may have its

neural correlate on different sets of neurons being tuned to different properties. For instance, when seeing the following sequence of lines \\\|||///, the reader is likely to group it into three sets, according to the orientations of the lines. This may be due to neuronal populations which respond to lines oriented at 135°, at 90° and at 45° being activated by these lines. These neuronal populations are likely to be distinct as the orientations of the lines are distinct enough. Thus, three groups may be perceived due to the activation of three distinct neuronal populations (Todorovic, 2011).

The Law of Common Fate may be rooted in a similar biological principle to the two described above. When elements move with the same (or very similar) directions across the visual field, they may activate a neuronal population of cells which is tuned to respond to that direction of movement (see section 2.2.1.1 for a description of such neural populations). Neurons activated in this neuronal population may excite each other through short and long range interactions, and their outputs may be directed to higher visual centers with larger receptive fields (such as V3 and V5, which are known to process global motion). Again, the activation of neurons with large receptive fields high up in the visual hierarchy may lead to a percept indicating that the individual elements belong to a group.

All of the postulates described above can be simplified to suggest that when there are similarities in a visual scene, the simultaneous firing and excitation of distinct populations of neurons could provide the neural basis for the process of grouping (Todorovic, 2011). This supports the very famous idea in neuroscience that neurons that fire together, wire together (Hebb, 1949), and implies that neurons fire together because they have a common reason to do so (e.g.: similar orientation preferences).

2.3.2 Gestalt Principles and Their Relation to Contour Integration

As described in section 2.3.1 Gestalt principles postulate how elements in a visual scenes might be organized to make sense of it. Contour integration, and the association field hypothesis (Field et al., 1993), also aim to provide an understanding of how visual elements may be organized by the visual system. Unlike for Gestalt psychology, when the field of contour integration emerged, a large pool of detailed physiological knowledge of the visual system already existed. Thus, influential theories of contour integration, such as the association field hypothesis, related biology to the perceptual organization of elements in a scene. However, much of the work on contour integration has been guided by the principles first outlined by (Wertheimer, 1923). In particular, the gestalt Law of Good Continuation, Law of Proximity and Law of Similarity are of relevance for contour integration.

The Law of Good Continuation states that elements which follow a smooth global trajectory are perceptually grouped into an object. In contour integration, contours are typically defined as co-linearly or co-circularly aligned edge elements. An example of a figure showing both, the Law of Good Continuation, and co-circularly aligned edge elements can be seen in figure 2.6, panel **A**. Several studies have formalized the Law of Good Continuation into geometrical requirements in a stimulus. The effects of curvature (Smits and Vos, 1987), and the number of changes in curvature

(i.e.: going from a concave to a convex shape) (Feldman and Singh, 2005) have been quantified in order to better understand the properties of stimuli which might give rise to good continuation. However, Gestalt psychologists, and early studied which sought to formalize Gestalt laws used an array of different elements in their stimuli (e.g.: Beck et al. (1989); Smits and Vos (1987)), thus making it difficult to make inferences on the mechanisms giving rise to good continuation, and to what can be defined as a contour.

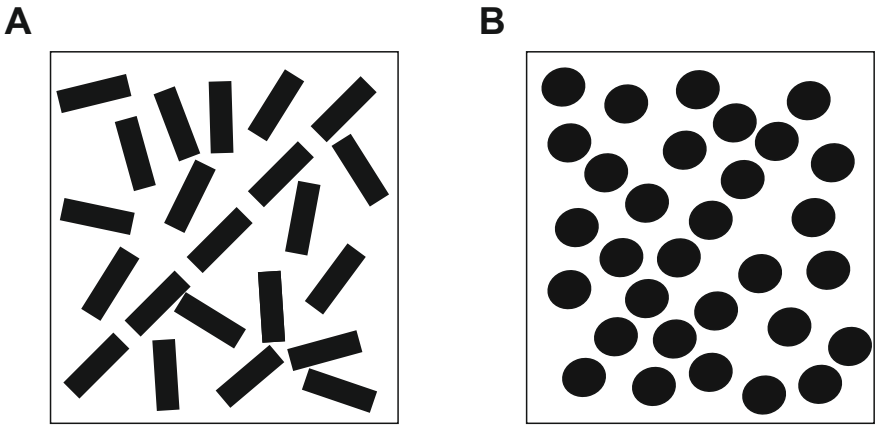

Figure 2.6: Examples of the Law of Good Continuation. **A**: Good continuation defined with edge elements with directionality; **B**: Good continuation defined with round elements with no directionality.

The stimuli shown in panels **A** and **B** of figure 2.6 can both be considered examples of good continuation. An observer most likely perceives a diagonal line expanding from the bottom left to the top right when viewing these stimuli. However, panel **A** is more likely to be regarded as a good example of a contour (as typically thought of in CI literature) than panel **B**, since in panel **A** edge elements with a direction are used to define the contour.

The Law of Proximity, which states that elements tends to be perceived as belonging to a group if they are in physical proximity to one-another, also plays an important role in contour integration. This Gestalt law has also been scrutinized in order to better characterize it. Rock and Brosgole (1964) constructed a rectangular grid in which rows of dots were separated by three inches, and columns by four inches. This induced the perception of the dots being organized into columns, however, if viewed from a control distance and rotated to approximately 41°, all dots had an equal retinal separation, which did not induce any sort of grouping. Thus, it appears that the Law of Proximity is not rooted in the physical characteristics of a stimulus, but rather in the manner in which stimuli are perceived and processed. Figure 2.7 shows an example of a stimulus which induces the grouping of elements in a scene based on their proximity (if viewed under the right conditions).

The association field hypothesis acknowledges the fact that it is not physical proximity, but rather retinotopic proximity, which induces the grouping of elements. Thus, the hypothesis states that edge detectors with *receptive fields* which are close to one another are more likely to share strong connections, than edge detectors with RFs which are not close to one another (Field et al., 1993). Electrophysiology (Shouval et al.,

2000) has provided evidence for this claim.

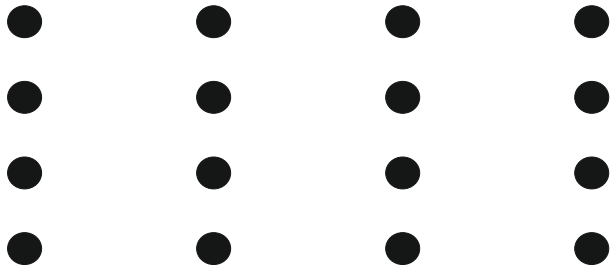

Figure 2.7: Examples of the Law of Proximity. If viewed under the right conditions the stimulus presented in this figure will lead observers to group the dots in the figure into columns rather than rows, because the dots have a smaller vertical separation than horizontal separation (i.e.: they are in closer proximity vertically than they are horizontally).

The Law of Similarity, which states that elements tend to be grouped together if they are similar to one another (their similarity can be conceived in may different ways, e.g.: by color, shape, spatial frequency, etc.) plays a role in CI when one thinks of this process as not only based on alignment cues. If contour integration is seen as a process that relies on the grouping of elements based on shared characteristics, then the role of similarity in CI becomes evident. McIlhagga and Mullen (1996); Persike et al. (2009) have conducted experiments under the umbrella of contour integration, in which elements in a stimulus were grouped based on cues unrelated to alignment. Figure 2.8 shows an example of elements in a visual stimulus which are perceptually grouped based on their similarity in terms of color.

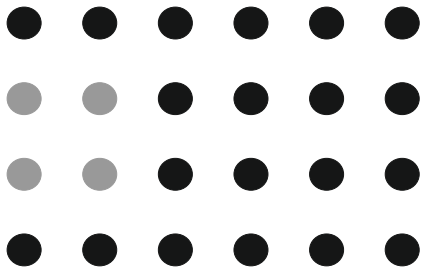

Figure 2.8: Examples of the Law of Similarity. The four grey circles located on the left of the stimulus are likely to be perceived as a group by an observer due to their similarity in contour, and dissimilarity from all other elements in this stimulus.

Since in contour integration studies, band pass band-pass elements (often gabor patches) are often used as edge elements to create contours, the spatial frequency of the elements is an obvious cue on which similarity can be established. Persike and Meinhardt (2015a) have shown that the spatial frequency of edge elements is a strong cue on which contours can be defined on the basis of similarity. Furthermore, Persike and Meinhardt (2016) also showed that spatial frequency similarity is a cue which can interact with alignment in order to support contour integration.

Thus, the grouping principles proposed by Gestalt psychologists in the early 1900's (Wertheimer, 1923) are closely related to the formal requirements for contour integration. In fact, Field et al. (1993) seem to have derived many of the ideas for the experiments which they performed in their seminal paper from these Gestalt principles. They probed observers ability to detect contours under conditions which manipulated good continuation, proximity, and similarity, amongst others. These experiments were some of the first to employ stimuli which allow for the methodical characterization of "the relationship between the tuning properties of single cells and the network operations describing how their outputs might be combined" (Wagemans, 2015). Although in most contemporary contour integration studies not much emphasis is given to the Gestalt principles outlined in this section, the field of CI heavily relies on them.

We have seen in this section that the process of feature integration is of out-most importance in helping us create our experience of the visual world. Although the process of creating this experience seems effortless, there are large amounts of computations which are required for features to be integrated into contours. Over the past century scientists have concerned themselves with identifying the processes, and later on the computations, which are required for us to create our experience of the visual world. This thesis also concerns itself with this topic. Now that the reader posses an understanding of the main grouping/integrating principles which allow us to perceive the world as we do (and their biological correlates), it is possible to understand the remaining chapters of this thesis.

Chapter 3

Contour Integration in Dynamic Scenes

3.1 Introduction

In a natural environment, the visual system receives a constant stream of dynamically changing and high-dimensional information which must be processed efficiently in order to create a coherent picture of our world. Throughout the past century several heuristics and mechanisms have been proposed to explain how the visual system undertakes such a complex computational task (see e.g. Gilbert and Li (2013); Heider (1970) for an overview and section 2.3.1). Contour integration (CI), the visual system's method of grouping discrete elements across the visual space into continuous contours, is thought to be one of the mechanisms which bridges the gap between primary sensory processing and object based perception (Li and Li, 2015). As such, this process has been extensively studied, however, it has almost exclusively been studied in static stimuli. Given that the natural visual environment is almost exclusively dynamic the question of whether the pop-out qualities of CI, as previously studied in static scenes, hold true for dynamic scenes. In two experiments the capabilities of observers to detect contours in dynamic scenes were evaluated. In the novel stimuli developed for these experiments Gabor elements rotated in a scene for a predetermined period of time before they aligned to generate a contour.

In this chapter the results of the experiments carried out with this novel paradigm will be reported. Of most interest, results revealed that in brief presentations (akin to static scenes) contour integration performance was relatively high (about 87%), however in extended presentations, when subjects were first exposed to a noise signal generated by rotating Gabors before they were presented with the same contour stimuli as in the brief presentation, performance dropped by about 20%. In order to achieve similar contour detection performance levels in a brief condition the presentation time had to be drastically reduced to (approximately) 50ms. Furthermore, a cueing paradigm revealed that cueing specific contour positions or shapes helps to partially restore contour detection capabilities, however, only in extended presentations is the combination of a position and shape cue more efficient than presenting either one of the cues alone.

The reader should note that the contents of this chapter parallels Grzymisch et al. (2017a), since the results of these experiments have already been published.

3.2 Contour Integration Background

The process of visual perception is an incremental one, and the process of organizing a visual scene into coherent percepts requires perceptual grouping. There are many ways to achieve perceptual grouping (i.e.: the combination of elements into meaningful configurations), and one of the most basic forms of grouping is described by contour integration (CI). Gestal psychologists were early proponents of heuristics and theories of visual perception which explained perceptual groping (for a review see (Wagemans et al., 2012a,b)). Their "Law of Good Continuation" (Coren and Girgus, 1980; Wertheimer, 1923), which states that visual elements following smooth global trajectories are perceptually grouped, is the basis of CI. Contour integration builds on this notion by adding the requirement of local oriented elements being aligned in a collinear or cocircular manner in order to be perceived as a global "contour".

Since contour integration is believed to be a basic visual process it has been extensively studied. Many studies have shown how contour perception is affected by the arrangement of local elements of a contour. Mandon and Kreiter (2005); May and Hess (2008); Strother and Kubovy (2006) have shown that contour detection performance deteriorates with increasing separation of contour elements, and Bex et al. (2001) has shown that contour detection performance also deteriorates if the local orientation of individual elements deviates from the global contour path. Others have shown that the physical properties of individual elements have effects in contour detection performance. When elements forming a contour are embedded in a background of randomly oriented elements, contour detection performance can be increased if the spatial frequency of contour elements is similar (and dissimilar to the background elements) (Dakin and Hess, 1998; Persike and Meinhardt, 2015b), or if contour elements have identical phases (which are distinct from the phases of background elements) (Hansen and Hess, 2006).

The multitude of variations under which contour integration has been shown to take place indicates that CI is a very robust process. Furthermore, neural correlates of contour integration have been shown to emerge regardless of behavioural requirements (Bauer and Heinze, 2002), and contour integration has been shown to occur in a very rapid time scale. Stimulus presentation times of just 30ms have been shown to be sufficient for observers to reach 75% correct performance in a 2AFC contour integration task (Ernst et al., 2012) and for macaques to reach a 66% performance in a contour discrimination task with a 25% chance level (Mandon and Kreiter, 2005).

The nature of the findings presented above, alongside with the proposed mechanisms for contour integration by Field et al. (1993) in their seminal paper suggest that contour integration has a pop-out nature. Thus, that feed-forward neural mechanisms, and/or recurrent integration of visual information are the underlays of CI, suggesting that CI is barely influenced by cortical feedback or current cognitive states. As such, early studies (Bauer and Heinze, 2002; Polat et al., 1998) on the neural correlates of contour integration focused on V1 as a likely source for the effects described by subjects when contour integration is thought to occurs (the effect being a strong percept of an object

embedded in a noisy background). However, it was soon discovered that horizontal connections in V1 cannot fully account for the contextual modulations effects seen in contour integration (Angelucci et al., 2002), thus, it has been suggested that higher order visual areas likely play a role in contour integration. Chen et al. (2014) has found evidence in support of this suggestion as they observed V4 to have a faster and stronger response to contour stimuli than V1. In a similar line of evidence, Gilad et al. (2013) found that responses to individual Gabor elements comprising a stimulus are first observed 40-140ms after stimulus onset, regardless of whether the elements are part of a contour or not. However, the increment of V1 firing rates due to the alignment of individual elements of a contour is seen approximately 150-250ms after stimulus onset Gilad et al. (2013), suggesting the involvement of higher order visual areas (as V1 circuits would typically respond faster if they were acting alone).

3.3 Motivation

Given that neurophysiological studies (Chen et al., 2014; Gilad et al., 2013) have found correlates for CI in areas other than V1 the assumption of CI not involving cortical feedback can be disputed. Furthermore, given that areas V2 and V4 are influenced by attention (Luck et al., 1997) (see sections 2.2.2 and 2.2.3), it is reasonable to assume that CI can be influenced by an observer's current cognitive state. However, in contour integration literature the question of how attentional states affect an observer's performance has rarely been addressed, and a key aspect of natural vision has also been largely ignored. Most CI paradigms use static stimuli, thus little is know about CI in situations which involve dynamic stimuli. Motion is a key component of natural vision, we have little trouble making sense of scenes with motion and we often use motion cues to organize the elements of a visual scene into a coherent percept. Given that this component of vision is so ubiquitous it seems evident that its effects should be studied alongside other basic visual processes such as CI. Although this might seem evident, most CI studies use static stimuli, the stimuli are often presented for a brief period of time (typically 200-300ms) and then masked. Hence, little is known about CI in situations when there is a smoothly changing stimulus which is presented for an extended period of time (e.g.: 2-3s).

The experiments reported in this chapter set out to investigate how well different forms of attention (spatial attention, feature attention, and a combination of spatial and feature based attention) can support contour integration under two distinct stimulus presentation conditions, a brief presentation condition (235ms) and an extended presentation condition (>1800ms). Considering that contour integration has been shown to be a robust process, it would be expected that contour detection performance under brief and long presentation conditions would not differ greatly. However, the perceptual experience of viewing a scene for a few hundred milliseconds is quite different than that of viewing a scene for a few seconds. Different cognitive processes could take place while viewing scenes under these two distinct timing conditions. In extended viewing conditions (over a few seconds) an observer might experience changes in cognitive/attentional states with competing expectations forming and being evaluated against the visual ev-

idence. However, in brief viewing conditions these cognitive/attentional states might not have sufficient time to engage. The experiments reported in this section aim to evaluate whether or not an observer's contour detection performance is affected by a timing manipulation (in which brief and extended viewing conditions are contrasted) while viewing a dynamic scene. This contrast is of relevance as much is known about the basic visual process of contour integration, however, most of what is known has been found out under short viewing conditions with almost exclusively no forms of motion. Considering that extended viewing conditions, with characteristics similar to those aforementioned, are the default situation in natural viewing conditions, it is desirable to study contour integration under extended viewing conditions as well.

3.4 Experiments

A field of edge elements realized as Gabor patches was used to conduct two contour integration studies. Observers viewed the stimulus under distinct timing conditions and were asked to report on the location of a contour. Manipulations aiming to target spatial attention, feature based attention, or a combination of these two attentional modalities were conducted. Experiment One and Experiment Two differed in the manner in which these attentional modalities were targeted.

3.4.1 Methods and Materials

3.4.1.1 Apparatus

Visual stimuli were displayed in a CRT monitor (Chuntex Electronics, Ultra Screen VL950T) with a refresh rate of 85 Hz and pixel resolution of 1152 x 864 (37.1 cm x 27.8 cm). The monitor was color calibrated using a gray scale correction performed with a gamma-corrected linear staircase consisting of 255 steps ranging from 0.11cd/m^2 to 113.1cd/m^2. Matlab (Mathworks, Inc.) was used to create the stimulus ensembles and to render them in real time with the Psychophysics Toolbox 3 (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997). The stimulus presentation area was limited to a quadratic aperture of 19.7°x 19.7° of visual angle (864 x 864 pixels). Observers viewed the screen binocularly at a distance of 80 cm in a room with attenuated light. Head movements were restricted using a combination of a chin rest and a forehead rest and an Eyelink II eyetracker (SR Research ltd.) was used to perform fixation control with a sampling rate of 250 Hz. Responses were provided with two custom-built response buttons with a temporal resolution of 0.1 ms.

3.4.1.2 Stimuli

Visual stimuli consisted of ensembles of oriented line elements realized as Gabor patches. Each Gabor patch in the ensembles rotated at a particular speed and with a particular direction determined by a random walk process (see "Stimulus dynamics" below for a detailed description). At a predefined time, from hereon forwards referred to as $t_{aligned}$, a subset of the Gabor patches in the display dubbed Gabors-in-contour (G_c) aligned

to form a contour which was the target of the subjects' visual search (see figure 3.1, panel **B** for an example of a stimulus with a contour). Each stimulus contained one contour of $G_c = 10$, and the contour was designed to be either right-oblique (45±10°) or left-oblique (135±10°), with a mainly straight global orientation, and was centered at one of the four quadrants of the stimulus display aperture. The locations and orientations of contours in the stimuli were equally balanced amongst all defining qualities of the target (i.e.: orientation and placement in the stimulus) and pseudorandomized for presentation.

Single element properties and placement: The position of individual Gabor patches in the display stimulus was determined by a procedure which created a hexagonal grid with N elements and then subjected the positions on the gird to a placement shifting process realized as a random walk with an additional constraint on minimal element distance of 0.21° of visual angle. Contours were embedded by selecting a set of $G_c = 10$ centered at a location close to the center of one of the four quadrants in a grid of positions and aligning them with a specific orientation (i.e.: left or right oblique). The orientation of individual elements in G_c at $t_{aligned}$ was set tangentially to a spline curve connecting the contour elements with minimal total curvature. The remaining elements in the display (Gabors-in-background or G_b) were given a random orientation at $t_{aligned}$. On average a stimulus consisted of 560 Gabor patches ($G_c = 10$ and $G_b = 550$) rendered with an even cosine phase and a spatial modulation period λ of (approximately) 0.11° of visual angle. The standard deviation σ of the Gaussian envelope was set to 0.138° of visual angle, and the average separation between neighbouring elements was 1.17° of visual angle.

Stimulus dynamics: The dynamic stimuli were presented for a time T during which elements rotated with different speeds and in different directions (clockwise or counterclockwise). A random walk process on the angular velocities $\omega_i(t)$ was used to create the individual rotation trajectory and speed for each of the elements i in a stimulus. The rotation trajectory is dubbed $\phi_i(t)$ and $\omega_i(t)$ was given a non-zero drift velocity ω_r:

$$\tau\dot{\omega}_i(t) = -\omega_i(t) + \omega_r + \sigma_r\mathcal{N}(0,1) \quad (3.1)$$

$$\dot{\phi}_i(t) = \omega_i(t) \quad (3.2)$$

The above equation was parametrized as follows: $\omega_r = 100°/\text{s}$, $\sigma_r = 120°/\text{s}$ and $\tau = 3.3$ s.

The rotation of elements close to the center of the display, thus in the subject's foveal region, was particularly salient, thus rotation speeds were scaled with element eccentricity e_i according to the following equation:

$$\omega_i^{scaled}(t) := \omega_i(t)\,(1 - \exp\,(-e_i/\lambda_e)) \quad (3.3)$$

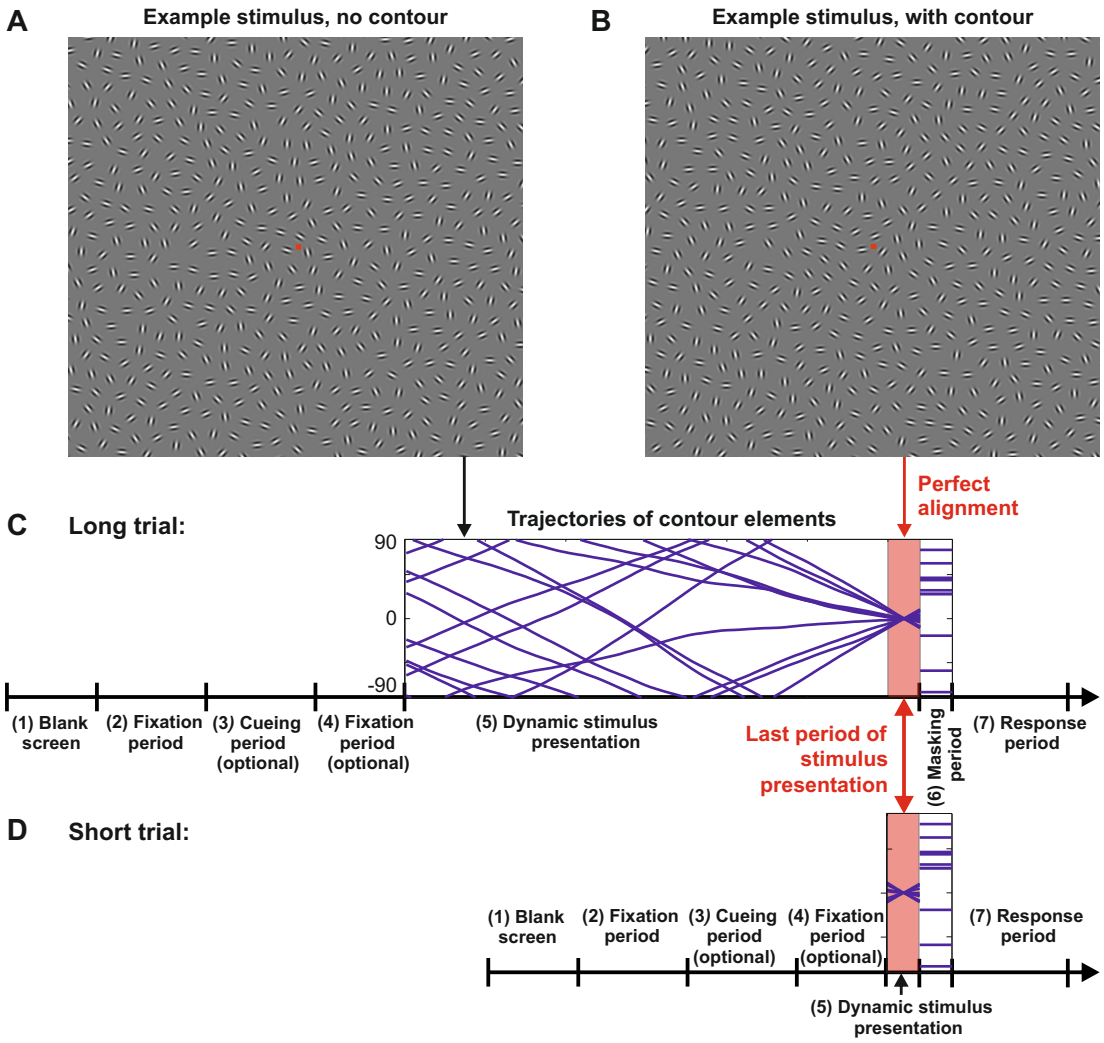

Figure 3.1: Example stimuli and trial sequences.(**A**) Example stimulus with a random arrangement of the Gabor patches. For most of the dynamic stimulus presentation time in the **Long** condition the rotating Gabors generated no meaningful figures, only towards the end of the stimulus presentation they aligned to form a contour as shown in example stimulus (**B**). Here, the contour can be seen on the upper right quadrant, tilted at 45°.(**C**) **Long** and (**D**) **Short** trial sequences. The same sequence of events was followed for **Long** and **Short** trials, they only differed in the stimulus presentation times (SOAs) (period **5**). The sequence of events for a trial was laid out as follows: (**1**) Blank screen for 1176ms; (**2**) Fixation spot for 2353ms; (**3**) Optional period – presentation of a cue in cued trials (see Methods); (**4**) Optional period – fixation spot for 588ms in cued trials (see Methods); (**5**) Stimulus presentation. The time of stimulus presentation varied depending on the test conditions and it is refer to as T in the text; (**6**) Masking period for 588ms; (**7**) Response period for 2353ms. For period **5** (stimulus presentation) either the **Peak** SOA ($T = 235$ms), an adjusted SOA for the **Short** condition ($T_{average} = 90$ms), or one of three different SOAs were used for the **Long** trials ($T = 1882$ms; $T = 2823$ms; or $T = 3764$ms). Presentation of these three distinct times in the **Long** condition were pseudorandomized. The perfect alignment period was identical for **Long** and **Short** trials as **Short** trials were realized by only presenting a section of a **Long** trial. Perfect alignment occurred shortly before the presentation of a mask (for **Long** trials $T_{aligned} = T - 117.6$ms and for **Short** trials $T_{aligned} = T/2$). Note that this figure was also presented in (Grzymisch et al., 2017a).

With $\lambda_e = 4.2°$, the rotation speed averaged approximately 20°/s near the fovea and 80°/s at more than 5° of visual angle eccentricity. After creating the trajectories for each individual element, the initial phases $\phi_i(0)$ for the G_c were shifted. This was done so that at a predefined time $T_{aligned}$ the contour elements would be aligned to the contour path. $T_{aligned}$ was always close to the end of a trial. More specifically, for trials with $T \geq 235$ ms, $T_{aligned}$ was $T - 117.5$ ms, and for $T < 235$ ms it was $T/2$. Given the choice of parameters, the maximal misalignment of the contour segments to the contour path was approximately ±10° during a 235 ms period. Given what is known about contour detection performance such a miss-alignment does not lead to great decreases in performance, thus, in presentations with $T \leq 235$ ms contours were at approximately their maximally visible alignment.

The value of σ_r for the random walk was chosen so that after rotations of 180°, the spread of individual trajectories would have a standard deviation of about 45° around this value, thus ensuring that perfect alignment of the 10 contour elements did not occur before the appearance of the target at $t_{aligned}$. Because of the chosen value for σ_r and because the orientation of all elements at $T = 0$ was randomly chosen in a 0 to 2π interval the chances of spurious contours of the same length as that of the target contours appearing in the stimulus were negligible.

For a video sample of the stimuli please refer to this hyperlink: Video samples in supplementary materials[1].

3.4.2 Procedure and Task

The experiment consisted of a contour detection task in which subjects were presented with dynamic stimuli and were required to report whether they had seen a contour on the left or the right hemi-field of the stimulus. Three different timing conditions (**Long**, **Peak** and **Short**) existed, and four cue variants: **No-cue**, **Shape-cue**, **Position-cue**, and **Combined-cue**. **Combined-cue** was a combination of the position and the shape cues. At the start of the experiment subjects received a standardized training.

Training: The training consisted of four stages, each stage consisting of 24 trials. The training was designed to increase in difficulty as the different stages progressed. In order to facilitate training all stages were done with the easiest of the three timing conditions (i.e.: the **Peak** timing in which $T = 235$ms). In the first stage of training subjects were presented with stimuli which consisted only of a contour, that is, a contour was presented on a gray background. This stage of training was intended to introduce subjects to the task and to psychophysics as many subjects had never participated in an experiment before. At this stage subjects were instructed to become familiar with the shapes and locations at which a contour could appear, and with the rhythm of the task.

The second stage of training consisted of a contour stimulus with background elements, here the mask was omitted to help subjects detect the contours. The third

[1] Video 1: Peak condition, with added markers indicating contour location; Video 2: Peak condition, no markers; Video 3: Long condition, with added markers indicating contour location; Video 4: Long condition, no markers. If the link cannot be accessed directly please refer to section the supplementary materials for the URL.

stage of training consisted of a trial as it would be seen in the **Peak** condition, however, without the added requirement of fixation control. Subjects were instructed to maintain fixation, however, as there was no fixation control trials were not aborted if subjects performed a saccade, explored the stimuli, or blinked throughout a trial.

The fourth and last stage of training consisted of 24 trials which were performed in the exact same manner as they would be performed in the experiment, and which consisted of stimuli and stimuli sequences which were the same as they would be in the experiment. After every stage of training subjects were encouraged to raise any questions or needs for clarification which they had. Whenever a new variation was introduced into the experiment (i.e.: a new timing condition or cue type) further training was provided. The further training consisted of eight trials of the new variant of the task as it would be performed in the experiment and it was provided immediately before the subjects performed the new variant of the task. As in previous stages of training subjects were asked to raise any questions or need for clarification before the experimental block started.

Task: Subjects were required to fixate at a central fixation square and perform a two alternative forced choice (2AFC) task in which they had to indicate, as fast as possible, after they believed they had seen a contour in the display, whether they had seen the contour on the right or the left hemi-field of the screen. In order to provide their responses subjects were equipped with response buttons which were held in each hand and pressed with the corresponding thumb. After each response subjects were presented with auditory feedback which indicated whether or not their response was correct. A low frequency tone was used for incorrect answers and a high frequency one for correct answers.

All trials in the experiment followed the time line outlined in figure 3.1. If subjects broke fixation during periods *2, 3, 4* or *5* by blinking or generating a saccade then the trial was aborted and rescheduled for later presentation.

A mask was used at the end of the dynamic stimulus presentation to prevent afterimages of the contour remaining on the retina (Bacon-Mace et al., 2005). The mask was realized by giving a random orientation to all elements in the display and holding that scene for a period of 588ms. Thus the mask not only served the purpose of preventing afterimages but it also served to indicate the end of a trial.

For clarification purposes please note that in the following sections the time T at which the dynamic stimulus was replaced by the static stimulus (i.e.: the mask) will be referred to as the stimulus onset asynchrony (SOA).

3.5 Experiment One and Two

The training procedures, trial progression and task for Experiment One and Experiment Two were identical. The two experiments differed only in their block designed and in the way in which cues were realized.

3.5.1 Experiment One

After observers were subjected to the training procedure described in section 3.4.2 they started the experiment by performing the contour integration task in the **Peak** timing condition ($T = 235$). We used the **Peak** timing condition to evaluate a subject's ability to detect contours in our task, previous studies (Braun, 1999) indicate that detection performance reaches a maximum at approximately $T = 200$, thus, with a presentation time of 235ms in this condition we expected subjects to be at the peak of their performance capabilities. The **Peak** condition was evaluated at the start of each experimental session (2 sessions over 2 days), and each experimental block of the **Peak** condition consisted of 96 trials, yielding a total of 192 trials over two days for this condition. If subjects were successful in this task, that is, if they reached a correct contour detection performance equal or greater than 60% then they were tested in the **Long** condition. The **Long** condition consisted of trials with three distinct dynamic stimulation times ($T = 1882$ ms, 2823 ms and 3764 ms)[2] which were presented in a random order. Each **Long** block consisted of of 96 trials, and the condition was repeated at the start of each experimental session over the two days, thus also yielding a total of 192 trials. In order to proceed with the experiment subject's performance was also required to be equal or higher than 60% in this task.

After the **Peak** and the **Long** condition were tested it was necessary to establish a performance baseline for different Ts in non-cued conditions so that performance gains due to cue manipulations could be contrasted under brief and **Long** dynamic stimulation times. One of the goals of these experiments was to explore the benefits for contour detection performance when information about the contour's position and/or shape was known by the observers a priori to the contour's appearance, and to determine if this knowledge interacts in anyway with the length of dynamic presentation times. Performance in the **Long** timing condition was chosen as a baseline as it was lower than performance in the **Peak** condition, thus a staircase procedure was developed in order to match subject's performance in a brief dynamic stimulus presentation time, dubbed the **Short** condition, to that which had been observed in the **Long** condition.

3.5.1.1 Staircase Procedure

An iterative Bayesian scheme was used in which a set of candidate frame numbers k spanning a range from k_{min} (minimal SOA $T = k_{min}/f$, where $f = 85$ Hz - the frame rate of the monitor) to k_{max} (maximum SOA $T = k_{max}/f$) was initially selected based on a subject's performance in the **Long** task. Starting from a uniform prior distribution the probability $P(k)$ that a frame number k would give the desired performance p_{Long} was iteratively computed. In each iteration a test frame number k_{test} was sampled from $P(k)$, a stimulus using the corresponding SOA was presented, and the observer's response to said stimulus was recorded. Taking into account the number of correct responses for each frame number, $P(k)$ was computed as the product of

[2]We originally designed stimuli to be shown at a refresh rate of 100 Hz, but decided to use 85 Hz instead. For this reason, presentation intervals often have "peculiar" values such as 1882 ms (this value would have been 1600 ms at 100 Hz).

a. the likelihood that the number of correct responses for *lower* frame numbers was obtained from performances *lower* than p_{Long}

b. the likelihood that the number of correct responses for *higher* frame numbers was obtained from performances *higher* than p_{Long}, and

c. the likelihood that the number of correct responses for k frames was obtained from a performance p_{Long}

under the constraint of a 2-AFC paradigm which implies underlying performances for all k's between 0.5 and 1 and using binomial statistics. As a final step $P(k)$ was normalized so that $\sum_k P(K) = 1$. After a fixed number of trials (96), the procedure described yielded an estimate $k^\star = \text{argmax}_k P(k)$ of the SOA $T_{Short} = k^\star / f$ which should lead to an observers performance which approximates p_{Long}.

A set of six candidate frame numbers k, starting from $k = 5$, $k = 7$, $k = 8$, or $k = 9$ for performance values of $p_{Long} < 0.65$, $0.65 \leq p_{Long} < 0.7$, $0.7 \leq p_{Long} < 0.75$, or $0.75 \leq p_{Long}$, respectively, was chosen to conduct the staircase procedure. The number of candidate frames was chosen based on previous experience on how performance depends stimulation times in a contour detection task (Ernst et al., 2012).

Finally, after successful completion of the staircase procedure observers were tested in the **Short** condition with their individual recommended SOA $T_{Short} < 235$ ms with $n = 96$ (the staircase procedure and the test with the recommended $T_{Short} < 235$ was repeated on each experimental session) to ensure that their performance was approximately the same as their performance in the **Long** condition. The recommended T_{Short} was employed for all further stages of the experiment with brief presentation times.

3.5.1.2 Block Design

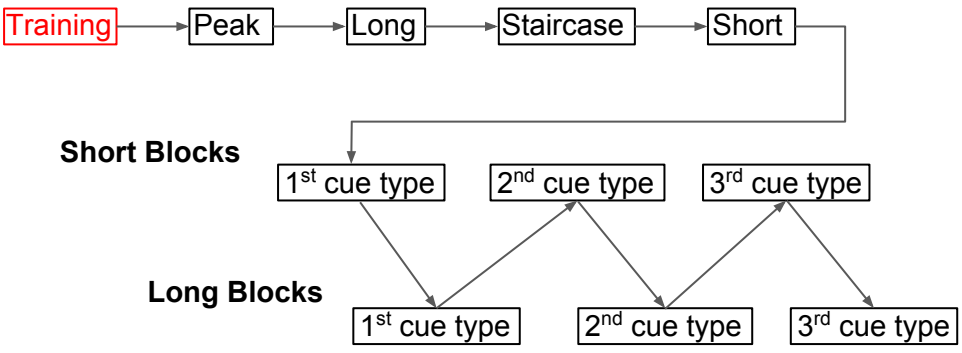

Figure 3.2: Time line for one experimental session of Experiment One. The initial step **Training** only took place in the first experimental session, and the order of **Long** and **Short** cued blocks, that is, with which one subjects started, was switched for the two experimental sessions. 1^{st}, 2^{nd} and 3^{rd} cue type refers to the **Position**, **Shape**, **Combination** cues and any of the 3 could be the 1st in this cyclical sequence.

The training and non-cued blocks in the experiment proceeded in the manner described above for all subjects. However, the cued blocks were presented with different schedules for each of the subjects. A pseudorandomized schedule of cued presentation was

generated for each subject prior to the start of the experimental session. In the pseudo-randomized schedule the three possible cues (see section 3.5.1.3) were cyclically rotated for each subject in the order of **Position**, **Shape** and **Combination**, so that if subject one started the cued blocks by performing his/hers first block with the **Position** cue, subject two would start by performing his/her first block with the **Shape** cue.
The block order for the two timing conditions in which cues were tested (i.e.: **Long** and **Short**) were alternated and done in the reverse order on the two experimental session. The pre-generated schedule also determined whether a subject would perform the **Long** or the **Short** block first, and again, if subject one started by performing the **Long** blocks first then subject two would start by performing the **Short** blocks fist. In essence, a typical experimental session followed the time line described in figure 3.2.

Each of the cued conditions consisted of 48 trials, and each condition was repeated twice (1 time in each of the two experimental sessions), making for a total of 96 trials for each subjects in each of the cued conditions.

3.5.1.3 Cues

In Experiment One the types of cues used were of an endogenous nature, that is, they conveyed information about the position and/or shape of the target contour in an indirect manner. As seen in figure 3.3 the **Position** cue consisted of a triangle which could point either up or down to indicate whether the contour would appear on the upper or lower hemifield, the **Shape** cue was an oriented bar which could either be right tilted or left tilted to indicate whether the contour would be left or right tilted, and the **Combination** cue was composed of an oriented bar and a triangle conveying the same information as the previously described cues. The **Position** cue always appeared above or below the position at which the central fixation square had been present, and **Shape** cue was always centered at the location where the fixation square had been present.

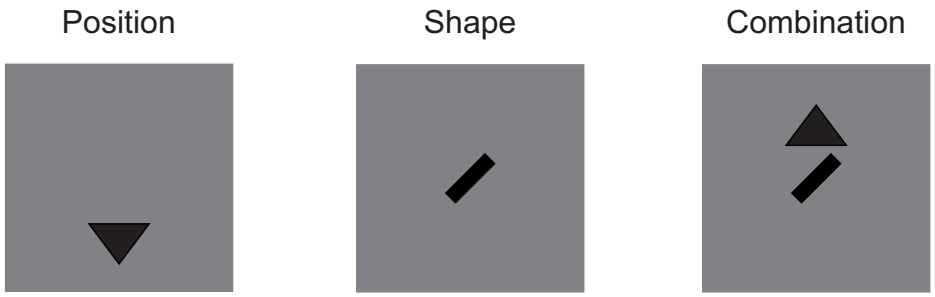

Figure 3.3: **Position**, **Shape** and **Combination** cues as labeled, for Experiment One. The cues were centered in the screen, in these images they have been enlarged to facilitate their visibility.

3.5.2 Experiment Two

Experiment Two differed from Experiment One in two main ways. First, in the way in which the cues were realized (see section 3.5.2.2), and second, in the block design

(see section 3.5.2.1). In Experiment One a pseudorandomized block designed was employed for the presentation of the cues, whereas in Experiment Two a fixed order for the presentation of cues was employed. As in Experiment One, when new subjects were invited for the experiment they first performed the same training procedure as described in section 3.4.2. Then they did the **Peak** condition with no cues to evaluate their ability to perform the task. If the participant was successful he then performed the **Long** condition, the staircase procedure, and then the cueing experiment.

It is important to note that for subjects who had participated in Experiment One and volunteered to participate in Experiment Two no training was given prior to the start of the cueing section of Experiment Two and they did not repeat the **Peak** condition, the staircase procedure, nor the blocks of 96 trials for the **Long** condition and **Short** condition. The SOA which had been determined for the presentation time in the **Short** condition of Experiment One was used to test participants in the **Short** condition of Experiment Two as well. Due to unforeseen perceptual learning effects (see section 3.6.3 for a detailed description) the performance of subjects in the **No-Cue Short** condition with the determined SOA for for **Short** condition in Experiment One did yield a matching performance to that seen in the **No-Cue Long** condition in Experiment Two. As such, subjects were invited to be re-tested in the **Short** timing condition after all data had been collected and analyzed.

3.5.2.1 Block Design

Subjects who had participated in Experiment One were already familiar with the task. The only change they experienced was the manner in which cues were presented. In order to familiarize subjects with the new way in which the cues were implemented a short training, consisting of four trials for each of the new cues, was provided prior to the commencement of the experiment. Subjects who had not participated in Experiment One went through the training procedure described for Experiment One, the **Long** condition, the staircase procedure, and the validation of the **Short** SOA (i.e.: after the staircase procedure was done a block of 96 trials with a **Short** SOA was done in order to gauge whether or not the subject's performance was approximately that observed in the **Long** condition) before starting the cueing blocks.

The cueing blocks were not pseudorandomized but rather fixed in order. The order of cue presentations employed was **Combination**,**Position**, **Shape** and **No-Cue**. Each block consisted of 48 trials, 12 trials for each of the cueing conditions. Two timing conditions were employed, the **Long** timing condition and the **Short** timing condition. In order to obtain 96 trials for each of the conditions the experimental blocks were repeated 8 times (for each for the timing conditions). Subjects were invited to come in on two consecutive days to perform the experiment, thus, they performed 4 repetitions of each block (for each of the timing conditions) on each day.

The presentation of the **Long** timing condition and the **Short** timing condition blocks was alternated, and every other subject started with a different timing condition. That is, if subject one start with the **Long** timing condition then subject two started with the **Short** timing condition. Furthermore, if a subject start with the **Long** timing condition on the first experimental session then they started with the **Short** timing

condition on the second experimental session.

The order of cue presentation was fixed to the **Combination**,**Position**, **Shape** and **No-Cue** sequence as it was hoped that by starting with the easiest of condition, and progressing in difficulty, subjects would become engaged in the task. Subject reports from Experiment One indicated that at times when subjects considered the task to be too difficult an were not successful in spotting the contours correctly they became disinterested/disengaged in the task.

3.5.2.2 Cues

The cues in Experiment Two were of an exogenous nature, that is, rather than conveying information through symbols as the cues of Experiment One did they conveyed information in a more direct manner (see figure 3.4).

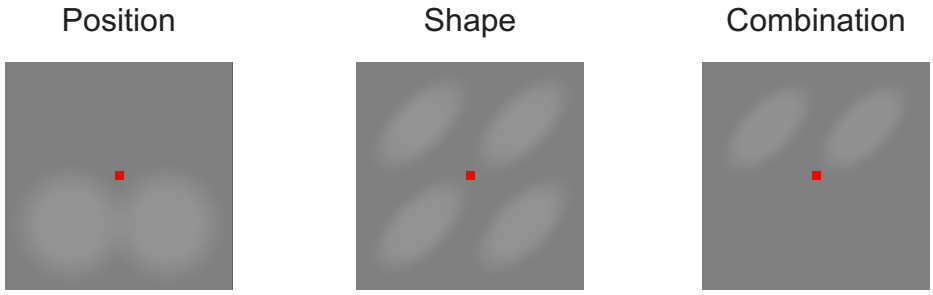

Figure 3.4: **Position**, **Shape** and **Combination** cues as labeled, for Experiment Two. The red square at the center of the panels is the fixation mark, the shadings are proportional to the shading size in the seen by the observers in the experiment.

The cues were redesigned for Experiment Two based on subject's reports of their experiences with the cues employed in Experiment One. Some subjects reported the appearance of the cues in Experiment One to be too abrupt and distracting, others reported the cues to be too salient, thus leading them to generate a saccade. In the case that subjects generated a saccade at this stage of a trial the trial would be aborted, as such their rhythm and flow in the experiment was often broken in Experiment One. The exogenous cues employed in Experiment Two succeeded in addressing the issued reported by the subjects as no subject reported problems with these cues.

3.5.3 Participants

10 participants (7 females) served as observers for Experiment One, however, it is worth to mention that a number of observers were invited for the experiment but needed to be dismissed at various stages before the start of the cued blocks because of their inability to perceive contours. Three subjects were unable to perceive contours in the **Peak** condition and four subjects were unable to perceive contours in the **Long** condition in Experiment One. In Experiment Two, 9 subjects had to be dismissed for their inability to do the task, six after the **Peak** condition and three after the **Long** condition. It was deemed that subjects were unable to perceive contours when their performance was

lower than 60% and by their own admission they reported no perception of a contour in the stimulus.

Of the initial 10 participants of Experiment One 9 participants volunteered to take part in Experiment Two. As noted in section 3.5.2 the **Short** condition in Experiment Two had to be repeated. Only 7 of the participants volunteered to be re-tested, hence new observers were recruited for this section. After all testing was done for Experiment Two the data set consisted of 9 participants (4 female), of which 7 participated in Experiment One as well.

All subjects reported normal, or corrected to normal vision, and provided written consent for the experiment in accordance to the rules and regulations of the University of Bremen local ethics committee. Subjects were paid 10 Euros per hour for their participation, and each experimental session lasted approximately 2.5 hours. The age range of subjects was between 24 and 41 years.

3.6 Results

Paired sample tests were used at all stages of the analysis with a significance level of $\alpha = 0.05$ unless otherwise stated.

3.6.1 Peak versus Long

In figure 3.5 correct contour detection performances and reaction times for the **Peak** and the **Long** conditions are compared. Both correct contour detection performance and reaction times were significantly different in the **Peak** and the **Long** conditions. Performance in the **Long** condition is 18.8% lower than in the **Peak** condition and significantly different ($t(8) = 10.3, p < 0.001$), implying that it is harder to detect contours in the **Long** condition than it is in the **Peak** condition. Reaction times point to the same observation as they were 268.8ms longer in the **Long** condition than in the **Peak** condition, and also significantly different ($t(8) = -9.6, p < 0.001$).

This is a surprising result for two main of reasons: (1) Because the 235ms period centered at $T_{aligned}$ of the **Long** condition, and the entirety of the **Peak** condition are identical. Thus, in the **Long** condition observers were exposed to the stimulus which lead to a performance of approximately 85%, however, by adding a history to this stimulus their performance dropped by about 18.8%; (2) because typically only about 100ms of exposure to a static stimuli of the kind we presented suffices to reach a performance of about 75%, and at about 200ms of exposure to said stimuli observers performance tends to be at a maximum (Braun, 1999). Thus, it seems that the **Long** condition (which is composed of a period of noise before the appearance of stimulus which lead to an 85% performance) was affected by the dynamic history which was presented prior to the appearance of the target, and that this history induced a suppression on subjects ability to detect contours.

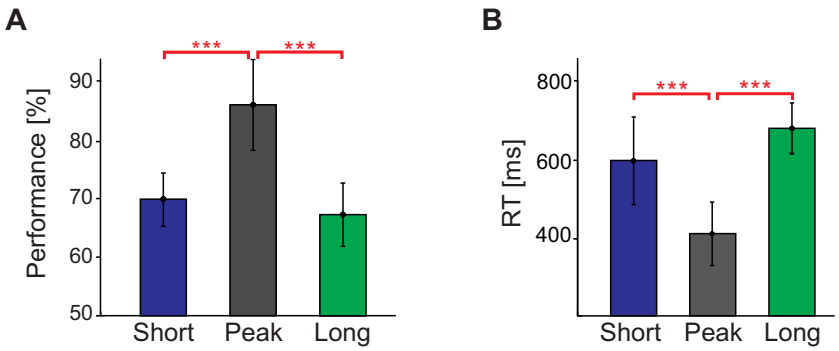

Figure 3.5: Comparison of different timing conditions without cues. Panel **(A)** shows performance and panel **(B)** reaction times for the **Short**, **Peak**, and **Long** timing conditions without cues. Vertical bars indicate standard errors. The **Short** data consists of trials performed with distinct Ts for each individual observer obtained with the staircase procedure in order to match their performance in the **Long** condition. Stars represent the level at which these conditions were found to be significantly different ($***p < 0.001$). Performances and RTs in the **Short** and **Long** conditions establish the baseline for the cueing experiments shown in the subsequent figures. Note that this figure was also presented in (Grzymisch et al., 2017a).

It is possible that the suppression in subject's ability to perceive contours in the **Long** condition develops slowly over time, and that the longer the exposure to a dynamic stimulus prior to the presentation of a target the lower the ability of a subject will be to detect contours. In the **Long** condition three distinct dynamic stimulation times were employed (1882ms, 2823ms, and 3764ms) in order to avoid expectancy effects from subjects, that is, in order to avoid subjects from developing an unconscious expectancy on when the contour would appear. Having these three distinct dynamic stimulation times served that purpose, however, it also allows for the hypothesis just stated to be tested. If performance is comparable between these three times then we can conclude that at least beyond the shortest dynamic stimulation time which we employed there is no decay in subject's ability to detect contours.

The analysis indicates that the performance for the three time intervals were almost identical (67.1%, 67.5%, and 67.5%) and that there were no statistical differences in performances amongst these three conditions ($F(2) = 0.01, p > 0.05$). The same result was obtained for reaction times (808.0ms, 774.9ms, and 745.9ms, respectively, and $F(2) = 2.99, p > 0.05$). This analysis indicates that the observed drop in performance between the **Peak** condition and the **Long** condition is due to a suppressive effect which takes place in a time scale which is shorter than 1.9s, and by that time it is fully developed.

This result, the observed dropped in performance between a brief exposure to a dynamic stimulus and a long exposure to a dynamic stimulus, is in stark contrast to our expectations and to the ease with which our visual system performs in dynamic visual

environments. One explanation for the phenomenon which we observed is that top-down processes, which are typically active in the process of vision when we experience dynamic visual environments, become important in long dynamic stimulation periods. Thus, these top-down processes which might aid subjects to focus on selected areas and/or features of interest in a scene could help compensate for the loss of subject's ability to detect contours in long dynamic stimulation periods.

3.6.2 Cueing Effects

With the introduced cued (which were designed to direct subject's focus to either a location in space, to a particular contour shape, or to both) the experiments allowed for the quantification of the extent to which the top-down processes realizing the focus on the features selected can improve contour detection. As the data has indicated that there is a suppressive effect in contour integration performance in long dynamic stimulation times it is also interesting to contrast possible performance gains given by a priori knowledge of certain aspects of the target of a visual search task in a long dynamic stimulation condition and a brief dynamic stimulation condition. In the brief dynamic stimulation condition the suppressive process(es) which led to the reduction in performance reported in section 3.6.1 should not occur, as this presentation time, by design, was shorter than the **Peak** condition. Thus, it is hypothesized that cues intended to aid the engagement of top-down processes, and thus lead to better contour integration performance, should be particularly helpful in long dynamic stimulation conditions and not so much so in a brief dynamic stimulation condition.

In order to test this hypothesis the **Short** condition was introduced (see section 3.5.1.1). On average, the time of dynamic stimulation employed in this condition was $T = 90ms$ ($S.D.$: $21.38ms$) around the point of perfect alignment. For Experiment One the **Short** and **Long** condition were not found to differ from each other (see figure 3.5), neither in performance ($t(8) = 1.2, p > 0.05$) nor in reaction times ($t(8) = 0.60, p > 0.05$), thus yielding a common ground on which cueing effects could be quantified in a long dynamic stimulation time ($T > 1.9$s) and a brief dynamic stimulation time ($T < 90$ms).

3.6.2.1 Exogenous versus Endogeneous Cues

In Experiment One and Experiment Two different cueing methods were used. The different cueing methods arose due to our experience with this newly developed paradigm. Results showed that subject's performance in the **Peak** condition was comparable to that seen in previous contour integration studies which employed static stimuli, however, in the **Long** condition observers performance was quite low and several observers needed to be dismissed from the experiment due to their inability to perceive contours. We reasoned that in such a difficult task cues needed to be presented in an optimal manner in order to effect significant improvements in performance. In the **Short** condition only two cues in Experiment One provided a significant improvement in performance when compared to the **No-Cue** condition. Only the **Shape** cue ($t(8) = 2.80$, $p < 0.05$) and the **Combination** cue ($t(8) = 3.02$, $p < 0.05$) yielded a performance which was significantly higher than that observed in the **No-Cue** condition, the **Position** cue

did not provide a statistically significant improvement ($t(8) = 1.06$, $p > 0.05$). In the **Long** condition only the **Combination** cued provided an improvement in performance ($t(8) = 2.6$, $p < 0.05$), the **Shape** and **Position** cues did not provide a statistically significant improvement in performance ($t(8) = 0.26$, $p > 0.05$ and $t(8) = 0.81$, $p > 0.05$, respectively). As such, the cues used in Experiment One, dubbed **Exogenous** cued did not prove to be very effective.

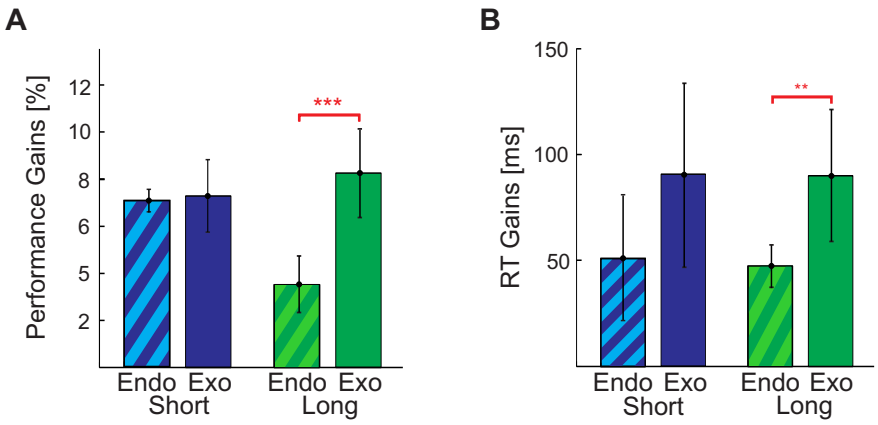

Figure 3.6: Comparison between **Endogenous** and **Exogenous** cues. Perceptual gains expressed as **(A)** increases in performance and **(B)** decreases in reaction times. Vertical bars indicate standard errors. Data was collapsed over all cueing conditions (**Position**, **Shape**, **Combination**) for **Endogenous** (candy stripes bars) and **Exogenous** cues (solid bars). Gains were calculated by subtracting the baseline condition (**No-Cue**) from the average over all cueing conditions. Differences between cue types in the **Short** timing condition were not significant. Significant differences at levels $***p < 0.001$ and $**p < 0.01$ were found for performance and reaction time gains, respectively, in the **Long** condition. Note that this figure was also presented in (Grzymisch et al., 2017a).

New cues were introduced in Experiment Two, these cues were dubbed **Endogenous** cues, and they provided reliable gains in performance when compared to the **No-Cue** condition (see figure 3.7). Furthermore, to test whether these two cueing methods led to different gains in performance the three types of cues were collapsed in both cueing methods and the improvements in performance and gains in reaction times yielded by these two cueing methods were contrasted with non-paired t-tests. Figure 3.6 shows this data and a clear pattern emerges for **Exogenous** cues providing greater gains in performance and larger reductions of reaction times for both, the **Long** and the **Short** conditions. Thus, **Exogenous** cues were more effective for our purposes, possibly because of the more explicit nature in which they conveyed the information regarding the properties of the target contour. When looking at panel **A** in figure 3.6 an exception is seen for the case of performance in the **Short** condition ($t(52) = 0.88$, $p > 0.05$), here the **Exogenous** cues did not lead to statistically significantly greater performance gains than the **Endogenous** cues. In the **Long** condition

the gain in performance and reduction in RTs for **Exogenous** cues was found to be significantly higher than that observed for **Endogenous** cues, $t(52) = 6.32$, $p < 0.001$ and $t(52) = 3.41$, $p < 0.01$, respectively. In panel **B** of figure 3.6 a significant effect for the difference in gains of reaction times for the **Short** condition could also be expected, however, this effect was not found ($t(52) = 0.56$, $p > 0.05$). A clear outlier is seen in the data for this condition, if removed the difference in gains between the direct and indirect cues for this condition almost doubles from 39ms to 71ms.

The results of the above mentioned analysis indicate that both types of cues (exogenous and endogenous) lead to improvements in performance and RTs, however, the **Endogenous** cues led to smaller perceptual gains than the **Exogenous** cues. Unsurprisingly the visual system seems to be better able to make use of the **Exogenous** cues than of the **Endogenous** cues, possibly because **Exogenous** cues are better able to engage top down processes. Since the **Exogenous** cues proved to be better suited for our purposes in the remaining sections of this chapter only results pertaining to the **Exogenous** cues will be discussed unless otherwise stated.

3.6.2.2 Effects of Single Cues and Cue Combinations

Two single cues (**Position** and **Shape** cues) were employed in the experiments, since these two types of cues were designed to drive attention to different aspects of a stimulus it was possible to combine the two cues. The two cues were combined in a new cue which is dubbed the **Combination** cue.

Each of the individual cues reduced the uncertainty space in the visual search task by a factor of two. The **Position** cue directly limited the area to which observers had to attend in order to detect a contour by half, and the **Shape** cue limited the number of tilt orientations to which subjects had to attend to by half. When the information of these two cues is added the uncertainty of the properties which a target contour can have is reduced to just two possible combinations, from an initial 8 combinations (when no cues are provided). As such, the combined cue was expected to yield the largest performance gains.

In order to test whether combining the **Position** and the **Shape** cue provided an added advantage when compared to the advantages provided by either of the single cues the data for the **Position** and the **Shape** cue were collapsed (colored bars in figure 3.7) and compared to the **Combination** cue. In the **Short** condition no differences were found between the averages of the single cues and the **Combination** cue ($t(8) = 1.15$, $p > 0.05$ and $t(8) = 1.15$, $p > 0.05$ for performance and reaction times respectively - see figure 3.7 panels **C** and **D**). In the **Long** condition both performance and reaction time gains were found to be greater for the **Combination** cue than for the mean of the individual cues (figure 3.7 panels **A** and **B**, $t(8) = 2.95, p < 0.05$ and $t(8) = 3.46, p < 0.01$ for performance and reaction times respectively). This indicates that there was a differential effect in how well observers were able to make use of individual cues or a combination of the cues in the **Long** and in the **Short** timing conditions, with the **Long** timing condition providing the necessary requirements for observers to make use of this combined information.

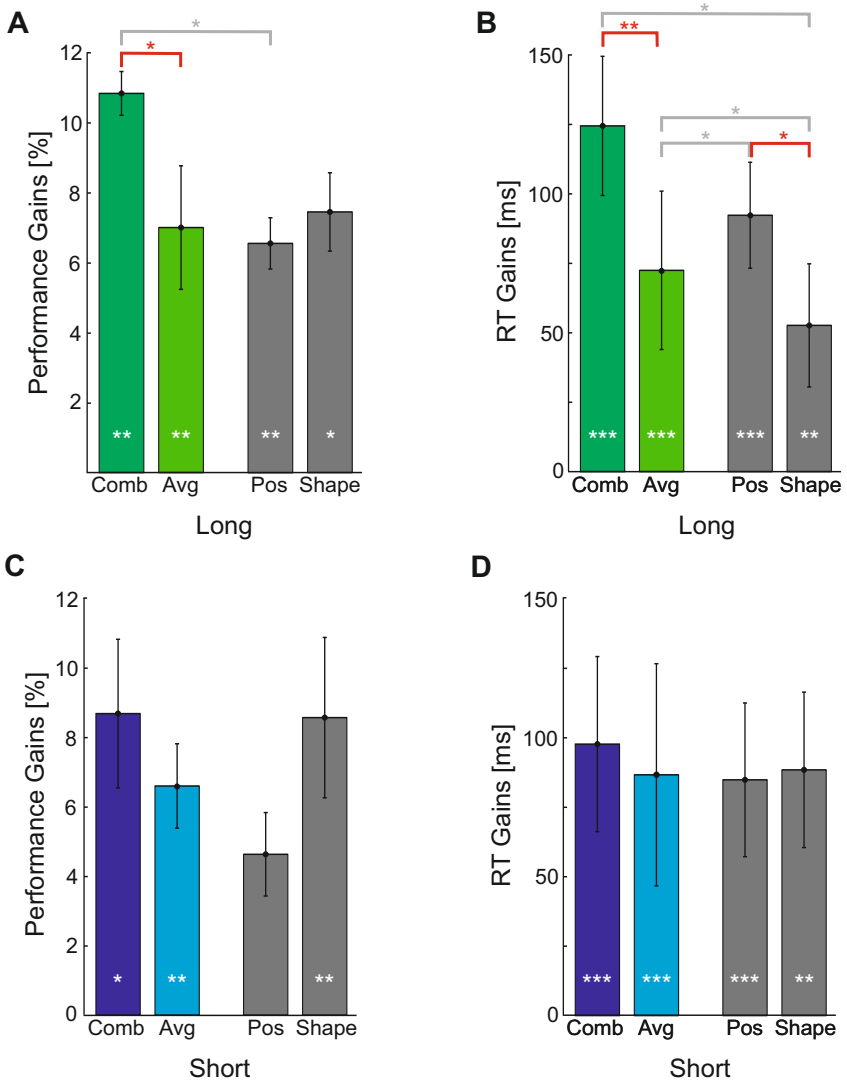

Figure 3.7: Perceptual gains provided by individual and combined cues. Perceptual gains expressed as **(A, C)** increases in performance and **(B, D)** decreases in reaction times, for the **Long** (panels **A, B** and **Short** timing conditions (panels **C, D**). Vertical bars indicate standard errors. Gains were calculated by subtracting the **No-Cue** baseline from the corresponding performances or RTs. Grey bars represent single cues (**Position** and **Shape**) whose average is displayed in the colored bar labeled *Avg*. White stars inside the bars represent the level at which the perceptual gain was significantly different from baseline. Stars above the horizontal bars represent the level at which two conditions were found to be significantly different (one star: $*p < 0.05$; two stars: $**p < 0.01$; three stars: $***p < 0.001$). Horizontal bars indicating comparisons discussed in the text are displayed in red color. Note that this figure was also presented in (Grzymisch et al., 2017a).

When comparing the effects that each of the individual cues had in the two timing conditions we see that in the **Long** condition the **Position** cue was more effective in reaction time gains when compared to the **Shape** cue ($t(8) = 2.52$, $p < 0.05$), but not in improving performance. In the **Short** condition the opposite pattern seems to

have taken place. No statistical differences were found between the perceptual gains provided by the two cues, however a pattern seems to have emerged. The **Shape** cue seems to have been more effective than the **Position** cue in terms of performance gains (with the **Shape** cue yielding approximately 4 percentage points more than the **Position** cue), however, no distinction between the two individual cues seems to have aroused in the reaction time gains.

When compared to the **No-Cue** condition both the **Position** and the **Shape** cues lead to improvements in performance and reaction times, for both, the **Long** and the **Short** timing conditions.

These results indicate that cues lead to performance gains, and that in the **Long** condition a combination of the **Shape** and **Position** cues had a special effect which is not accounted for by either of the individual cues. They also indicate that the different types of cues might act in a different way in short and extended dynamic stimulation times, however, due to the degree of variability between observers and sessions the data is not robust enough to make conclusions in this regard.

3.6.3 Perceptual Learning Effects

Perceptual learning is a common phenomenon in psychophysical tasks, as such it has also been reported in contour integration studies (Li et al., 2008; Schoups et al., 2001). In order to determine whether or not perceptual learning had occurred in these experiments performance data over the different experimental sessions, taking place in a chronological order, were examined. In order to better gauge any possible perceptual learning effects the data of Experiment One (**Endogenous** cues) is also presented here. The chronological order of data presented in this section is the following: **Endogenous** cues - first day, **Endogenous** cues - second day, **Exogenous** cues - first day, and **Exogenous** cues - second day, (or Experiment One, first day; Experiment One, second day; Experiment Two, first day; Experiment Two, second day). Only data for the **No-Cue** condition in the **Long** and the **Short** timings conditions was analyzed.

Observers' performance increased to a large extent over all experimental sessions, in the **Short** condition there was an increase of 16.8% from the first experimental session to the last experimental session. This increase was found to be statistically significant with a paired t-test ($t(8) = 6.74$, $p < 0.001$). There was also an increase in performance the **Long** condition from the first to the last experimental sessions. In the **Long** condition the increase in performance was not as large (only 10.5%), however, it was also found to be statistically significant with a paired samples t-test, $t(8) = 2.41$, $p < 0.05$.

The different amounts of perceptual learning which occurred in the **Short** and the **Long** condition were very consequential due to the design of this experiments. After a staircase procedure was done at the beginning of the first experimental of Experiment One it was ensured that performance did not differ in a statistically significant manner between the **Long** and the **Short No-Cue** conditions, as these were to be used as baselines. Indeed, at the end of the first session performance in the **Long** and **Short** conditions were found to be approximately equal (64.6% and 66.1% for **Long** and **Short** timing conditions, respectively), and a paired samples t-test did not reveal

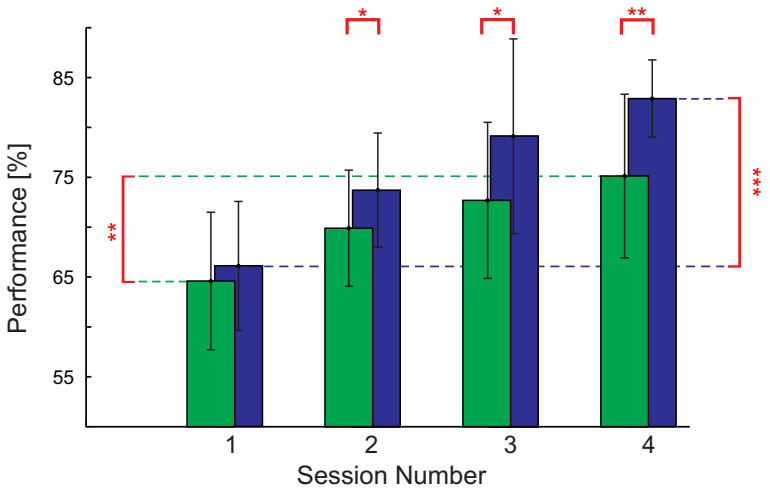

Figure 3.8: Perceptual learning. Contour detection performance and standard deviation (vertical bars) for the **Short** and **Long** timing conditions without cues, plotted over the subsequent four experimental sessions. Stars represent the level at which experimental sessions were found to differ from each other ($*p < 0.05$; $**p < 0.01$; $***p < 0.001$). Note that: (1) the actual time between experimental sessions varied among observers; (2) this figure was also presented in (Grzymisch et al., 2017a).

these two conditions to be distinct at an α level of 0.5. However, it seems that due to perceptual learning effects, on the second session performance between the **Long** and the **Short No-Cue** condition there was a 2.3% performance difference between these two conditions. With the **Long** condition having experienced an increase of 5.3% and the **Short** condition an increase of 7.6%, when compared to the respective performances for the **Long** and **Short** conditions in the previous session. Although the 2.3% difference in performance between the **Long** and the **Short** conditions seems small it was enough to lead to a statistically significant difference between these two conditions, $t(8) = 2.85$, $p < 0.005$. As can be seen in figure 3.8 there was a difference in the rate of improvement in the **Long** and the **Short** conditions over the four testing session, ultimately this different rate of improvement in the two conditions lead to a 7.7% difference between the **Long** and the **Short** timing conditions in the last experimental session.

3.7 Conclusion

The results obtained from these experiments provide a new line of evidence in the field of contour integration. Dynamic scenes were used to test contour integration in a previously untested scenario, the stimuli used did not consist of static images but rather used a dynamic process to generate contours. This dynamic processes allowed for the testing of contour integration in situations when the visual system is presented with a target stimulus which arises due to the changes occurring in the stimuli. Results indicated that when two identical stimulus are presented there is a large performance

difference when one of the stimuli is precluded by a dynamic history. In the **Peak** timing condition, which had a dynamic stimulation time of 235ms performance was approximately 85%, however, in the **Long** timing condition performance was approximately 65%. Thus, it seems that in the **Long** timing condition, after about 1.9s of dynamic stimulus presentation, the visual system was in a state that severely impaired the process of contour integration. In the **Peak** and **Long** timing conditions the last 235ms of stimulus presentation were exactly identical, and in this time the contour did not deviate from its perfect alignment more than $\pm 10^{\circ}$. Hence, it seems that there is no explanation for the stark drop in performance observed in the **Long** timing condition (when compared to the **Peak**) timing condition, other than the history which was presented prior to the last 235ms of dynamic stimulus presentation.

The results from the cueing experiments show that when observers' focus is directed towards certain features of the visual scene, which reduce the uncertainty of their visual search task, their performance increases, thus compensating for the loss in performance caused by the dynamic history of a target stimulus. In order to ensure that the effects of the cues were not unique to the **Long** timing condition the **Short** timing condition was introduced. The **Short** timing condition was designed to deliver a statistically similar performance in a brief presentation time (i.e.: less than 235ms) to that observed in the **Long** timing condition. This was desired as this condition would allow for testing of whether a loss in attention (or other effects which might lead to the drop in performance in the **Long** timing condition) could be uniquely revitalized by cues designed to promote the re-engagement of top-down processes. That is, whether the cues would have a greater impact in the **Long** timing condition, when the mechanisms which suppressed the perception of contours were already engaged, than in a condition which was not long enough for those mechanisms to engage.

Similar improvements in performance were found when cues were provided in the **Long** and **Short** timing condition. Thus, it seems that independently of stimulus presentation time, attending to particular features can improve contour detection. However, in the **Long** timing condition the visual system was better able to make use of the **Combination** cue.

A perceptual learning effect which lasted for weeks was also revealed by the data. As discussed in section 3.6.3, and as seen in figure 3.8, there was an increase in performance at every succeeding testing session. The fact that perceptual learning occurred, and that it lasted several weeks[3], suggests that functional changes leading to better contour integration performance must have occurred in observers' visual systems. As seen in figure 3.8 there was a differential effect of perceptual learning in the **Long** and the **Short** condition, with the **Short** condition being affected to a greater extent by perceptual learning.

[3]The testing sessions for Experiment One and Experiment Two were not conducted one immediately after the other. It was decided to conduct a second experiment, Experiment Two, after evaluating the data of Experiment One, hence some subjects were invited to come back for Experiment Two several weeks after they had participated in Experiment One.

3.8 Discussion

The reduction in contour integration performance seen in the **Long** timing condition, when compared to the **Peak** condition, was surprising since contour integration typically increases monotonically with increasing stimulus presentation times. Furthermore, it has been shown that with static stimuli a very short stimulus presentation times of about 60ms is required for macaque monkeys to reliably performance a contour integration task (Mandon and Kreiter, 2005), and that humans can also reliably perform contour integration with only 100ms stimulus presentation time (Ernst et al., 2012). In both the **Peak** and **Long** timing conditions there was a window of time of over 100ms in which the contours were almost perfectly aligned. Since the maximal rotation speed of elements at more than 5°of visual angle eccentricity was 80°/s the deviation in alignment of contour elements never exceeded more than ±10°in the **Peak** timing condition, or in the last 235ms of the **Long** timing condition. Since such a small jitter barely affects contour integration performance (Field et al., 1993) one can assume that with stimuli presented in the last 235ms of the **Long** condition subjects should have performed at their ceiling capacity. In the **Peak** condition this seems to have been the case, subject's performance was quite high, at about 85% correct. However, in the **Long** condition, although the exact same stimuli were presented in the last 235ms of dynamic stimulus presentation subjects experienced a stark impairment in their contour integration capabilities since their performance was at about 65%. Thus it seems that the visual system was in a state which impaired contour detection performance when the last 235ms of dynamic stimulus presentation were reached in the **Long** condition. Given that the natural visual world is very dynamic the results of this experiment may indicate that contour integration might have less importance than previously thought for visual perception, and that the process of contour integration, as observed in lab situations, may require a stronger support by parallel visual integration processes than previously thought.

The effects of dynamics in the stimuli employed in this experiment even lead to some subjects reporting no conscious contour perception in the **Long** condition. Given that no conscious perception of contours was reported by some subjects there might be implications for cognitive processes, such as perceptual learning, in the **Long** condition.

As the experiments conducted did not involve any sort of physiological measures it is only possible to speculate on the biological reasons behind the effects observed. Some of the possible reasons for the drop in performance observed in the **Long** timing condition might be the following: (1) processes of neural adaptation; (2) perceptual hysteresis; (3) attentional fatigue and large attentional requirements in the **Long** condition; (4) stimulus onset effects leading to transiently enhanced neural activity; and/or (5) temporal and recurrent processes in visual perception. To elaborate:

(1) Adaptation is a common phenomenon in neural systems. Many have described adaptation processes in the past, Kohn (2007) described how adaptation currents in neurons lead to decreasing firing rates when presented with a constant input current, Carandini et al. (1997) argued that recurrent inhibition leads to a normalization of neural responses, and Tsodyks et al. (1998) showed that synaptic resources deplete over time and that this depletion leads to neural adaptation.

Although the stimuli employed in these experiments were dynamic, Gabor elements did not move throughout the display, thus the same spatial receptive fields received input for a number of seconds in the **Long** condition. Stimulating the same receptive field for such a prolonged time may have led to adaptation in retinal cells on in the LGN, this adaptation would have in turn affected visual areas which are responsible for contour integration processes.

It is also possible that adaptation may have taken place at a cortical level. There are cellular mechanisms acting in cortical neurons which are responsible for adaptation processes after exposures to visual stimuli lasting several seconds. Carandini (2000) has shown that these processes can result in impairments in the perception of subsequent stimuli.

Motion adaptation has been well documented and it has been shown to impair perception. The effects of motion adaptation are often attributed to a reduction in the responsiveness of cells tuned to specific aspects of a scene (Anstis et al., 1998). In the **Long** condition the constant rotation of the edge elements in the display may have led to decreasing firing rates, thus, to a decrease ability to detect the appearance of a contour at the end of a trials. However, when considering this argument one must keep in mind that the rotation speed of the edge elements were slow compared to the time required for adaptation to occur. Reasonably assuming that a full rotation of a Gabor in the display took approximately 500ms one can expect that orientation selective neural populations in V1 would have recovered when a Gabor was once again aligned to the preferred orientation of the neural population in question.

The observation that top-down processes improve performance is in line with previous electrophysiological findings, Galashan et al. (2013) found that firing rates increase with increased levels of attention even if the neurons are in a sustained state of low activity due to adaptation effects. Previous pyschophysical results showing that attention increases the perceived duration of a stimulus and at the same time it decreases temporal resolution (Yeshurun and Levy, 2003; Yeshurun and Marom, 2008) are in line with Galashan et al. (2013)'s observations. Similarly, our cueing results are supported by Galashan et al. (2013)'s observations.

(2) The tendency of the visual system to stabilize a percept, dubbed perceptual hysteresis (Schwiedrzik et al., 2011), may be another possible explanation for the low performance observed in the **Long** dynamic stimulation condition. Keeping in mind that perceptual hysteresis is different from neural adaptation helps to separate the explanation presented above from this one. When human observers were scanned in MRI while experiencing hysteresis it was found that their ventral visual areas, superior parietal areas, and their frontal cortices had a high degree of activation (Kleinschmidt et al., 2002; Schwiedrzik et al., 2012). However, when subjects experienced adaptation in a similar task their visual areas V2 and V3 had increases in activity (Schwiedrzik et al., 2012).

In the **Long** condition of the experiment the "wrong" percept may have been stabilized, that is, a percept which was not the target contour may have been

stabilized and lead to a wrong response. Throughout the time of stimulation in the **Long** condition a number of percepts could have been perceived by observers, these could include proto-contours (short contours which may have been generated by the accidental alignment of two or three Gabors), or configurations of Gabor elements with regularities which may have lead to them being salient (e.g: star-like pattern). In the **Peak** condition it was unlikely that percepts other than the target contour would have appeared, or that they would have been more salient than the target contour. Furthermore, in the **Peak** condition perceptual hysteresis is unlikely to have occurred since 235ms is too short of a time for stable percepts to emerge.

(3) The attentional requirements to successfully spot the target contour in the **Long** timing condition may have been higher than the attentional requirements in the **Peak** or **Short** timing condition. In order to successfully spot the contour in the **Long** timing condition subjects had to attend to the stimulus for several seconds, whereas in the **Peak** timing condition the target contour was present in the stimulus from the start of the dynamic stimulation period.

Performance in visual search tasks has been shown to be dependent on the number of features to which an observer must attend in order to successfully perform the task and on the rate of stimulus presentation, with a higher number of feature which need attending to leading to lower performance, and with higher rates of stimulus presentation also leading to lower performance Fisher (1984); Joseph et al. (1997). The stimulus presentation times in these experiments (**Long** and **Peak**) could be analogous to the stimulus presentation rate in Joseph et al. (1997). If stimulus presentation times act in an analogous manner to stimulus presentation rate then this would explain the lower performance in the **Long** condition.

It can be assumed that attentional load was higher in the **Long** timing condition than in the other timing conditions simply because of the longer time span over which subjects had to attend to the stimulus to successfully spot a contour. Thus, the **Long** condition is more likely to have led to fatigue than the other timing conditions. Fatigue would have likely led to a narrowing of the spatial focus of attention, thus subjects may have resorted to shifting their focus of attention to different areas in the stimulus in order to spot the target contour. The shifting focus of attention could have been a systematic one, or it could have been driven by salient features in the stimulus, which could have been the target contour, but could have also been other structures (e.g.: short protocontours, star like patterns, etc.). As a result of the shifting of the narrow field of attention subjects may have missed the appearance of the contour since at the time during which the contour was presented their focus of attention may have been in another area of the stimulus. As the **Peak** condition was quite short, only 235ms, it is unlikely that subject would have shifted their focus of attention to different parts of the stimulus in this condition. Thus, the possible shifting of subject's spatial attentional focus, possibly caused by fatigue effects, would only have hindered performance in the **Long** condition.

(4) When there is an abrupt stimulus onset in which the visual scene goes from a blank screen to a screen with a large number of elements a transient in neural activity occurs (Jonides and Yantis, 1988; Visscher et al., 2003). This transient causes a decrease in the Fano factor (Galashan et al., 2013), thus implying that there is an increase in the signal-to-noise ratio, therefore, the system has better information processing capabilities. The transient effect described has been shown to have implication for contour integration. For example,Bauer and Heinze (2002) found that the initial transient of a V1 neuron's response is higher if the stimulus element inside its receptive field is part of a contour.

In the **Peak** timing condition the process of contour integration may have been enhanced by the transient in activity which occurred due to the switch of a blank screen to a screen containing a visual field full of edge elements. In the **Long** timing condition the process of contour integration would not have profited from the transient cause by the switch because by the time the contour appeared the transient would have likely extinguished. The type of transients described usually only last 50-100ms (Galashan et al., 2013) and in the **Long** timing condition contours appeared only after 1.7s or more after stimulus onset.

(5) Forward masking has been shown to affect a large number of perceptual tasks, change detection (Wutz and Melcher, 2013) and feature integration (Herzog et al., 2001) are of particular interest in this discussion. The dynamics of the stimulus prior to the presentation of the contour in the **Long** condition can be considered to be a forwards mask. If these dynamics are viewed as such, then the evidence provided by Wutz and Melcher (2013) would help explain why subjects may have missed the appearance of the contour. This was a meaningful change that seems to have been impaired by the the dynamics prior to the appearance of the contour (i.e.: the forwards mask). It is also possible that feature integration may have been impaired by the forward mask, as shown by Herzog et al. (2003).

Another option is to think of the dynamic stimulus prior to the presentation of the contour as a noise signal. This noise signal could have been integrated over time Burr and Santoro (2001); Drewes et al. (2015); Melcher et al. (2004). Burr and Santoro (2001) showed that the time scale of noise integration is approximately 2-3s, and that noise integration in this time scale leads to a plateau in performance after this time. In the **Long** condition a noise signal was presented for at least 1.9 seconds before a contour was presented. Three different SOA were employed in the **Long** condition (1882ms, 2823ms, and 3764ms), and as reported in section 3.6.1 no difference was found in the performance, nor in the reaction times, yielded by these three different SOAs. This result may be due to the effects of the noise signal presented prior to the presentation of the target being saturated, as shown by Burr and Santoro (2001).

The novel paradigm, which leads to the formation of contours by means of rotation of the edge elements in the display, was introduced in 2013 by a preliminary set of these experiments (Grzymisch et al., 2013). An EEG experiment (Castellano et al., 2014) employing a very similar paradigm indicates that in extended viewing conditions,

similar in length to the **Long** condition of these experiments, neural activity related to the appearance of a contour emerges very slowly. Castellano et al. (2014) found that neural signatures of contour integration only become significantly different from activity related to stimuli which do not contain a contour after approximately 150-250ms of the appearance of the contour, and need about 400 to 600ms to reach full development. Since in the paradigm used in this experiment the contour was presented for 235ms and then masked the slow processes which Castellano et al. (2014) cites would not have been able to fully develop and thus induce the maximum possible saliency of a contour. Thus, this can be taken as evidence for the contour integration process in the **Long** condition, or in extended presentations in general, to be dominated and perhaps suppressed, by ongoing processes in the visual system.

Another key finding from these experiments which suggests the involvement of higher visual areas than just V1 in the process of contour integration is the observed perceptual learning. Others (Gilbert et al., 2009; Li et al., 2008) have also observed perceptual learning in contour integration tasks, thus, the perceptual learning findings from these experiments are not unique. Typically perceptual learning is found in contour integration tasks because observers' visual systems are equipped to perceive contours in the manner that they are presented to them in an experiment, however, they are not used to doing so. But given that all the mechanisms required to perceive contours (in the manner in which they are presented in experiments) are present in observers' visual system they can very rapidly become accustom to seeing these new types of contours, and their visual system can learn to to employ the already existing mechanisms and substrates it posses to perceive these contours.

Contours presented in psychophysical tasks differ from those usually seen in nature because normally the stimuli used in a contour integration psychophysics experiment are design in a way which allows for the control of all parameters, and ensures that contours are only detectable based on the features of choice of the experimenter. In the case of these experiments the only feature which could lead to the perception of a contour was alignment. However, in natural images a multitude of different features (e.g.: alignment, color, texture, spatial frequency, etc.) may combine to indicate whether an element in the visual scene belongs to a given figure or object.

What is unique in the perceptual learning observed in these experiments in the differences in performance gains, given by perceptual learning, in the **Long** and in the **Short** timing conditions (see figure 3.8). This difference can be due to two distinct possibilities. One, the ceiling performance for the **Long** and the **Short** conditions could be identical, but perceptual learning rates in these could conditions could be are different, with the **Short** condition having a faster learning rate. Two, the perceptual learning rates for the **Long** and the **Short** condition are identical, but the two conditions have a different ceiling performance. Since we do not have data which shows the ceiling performance for these two conditions neither of the two possibilities can be taken as the ground truth. However, a reason to support the first hypothesis, that is, that perceptual learning rates are different in the two timing conditions is that in the **Long** condition contours were not always consciously perceived. If perceptual learning needs reinforcement then in a situation of uncertainty, such as that experienced by some subjects in the **Long** condition when they did not consciously perceive contours,

there would be no error signal that the visual system could employ to improve stimulus processing, thus leading to a slower learning process.

In summary, the results from these experiments suggest that in order to achieve a high contour integration performance in dynamic scenes support from top-down processes is needed. This finding is another piece in the puzzle of contour integration, it is in agreement with studies which show that top-down processes from higher visual or cortical areas support the process of contour integration (Li et al., 2008; Mijovic et al., 2014). Given that in much of the literature contour integration has been suggested to be a robust process which occurs mainly in the early visual cortex (Hess et al., 2003; Li, 1998; VanRullen et al., 2001) the stark drop in performance from the **Peak** condition to the **Long** condition was a surprising result. If it were the case that the process of contour integration is dominated by the early visual cortex then the current behavioural state of an observer should not bias results to a great extent.

Although new behavioural results have been provided by these experiments the experiments can not lead to conclusions regarding the source of the findings. It could be that attention, or other top-down processes, are independent of contour integration and simply act as an amplifier for the resulting neural states of the contour integration process in early visual areas. However, it could also be that top-down processes interact with contour integration processes happening in the early visual cortex and actively enhance the linking process of visual elements. The perceptual learning which occurred in these experiments provides good evidence for the suggestion that the process of contour integration does not solely occur in the early visual cortex. This claim is made since the perceptual learning which occurred in these experiments is different from that usually seen in V1. In V1 perceptual learning requires large amounts of repetitions of a task (2000-5000 trials (Schoups et al., 2001)) and is typically location specific (Schoups et al., 2001; Shiu et al., 1992). However, in these experiments perceptual learning occurred although the location in which contours were presented varied significantly (four quadrants and two orientations, with some freedom in the exact position where a contour could appear in the quadrant), and the number of trials to which each subject was exposed (about 200 at each quadrant with each of the orientations) was much smaller than the suggested 2000-5000 trials required for perceptual learning to occur in V1.

Thus, it can be speculated, but not definitely concluded that top-down processes invoked by cues which make observers focus on certain aspects of a visual scene always play a role in contour integration when a task is difficult, only when a contour is in a pop-out configuration it appears that bottom-up processes are enough for efficient contour integration to occur.

Chapter 4

Feature Integration on Alignment and Spatial Frequency Similarities

4.1 Introduction

Feature integration is an important process for visual perception. It is the piecing together of segments in a visual scene in order to extract coherent shapes/forms out of it. It can be based on distinct features such as co-alignment of edge elements, grouping of elements with similar properties such as spatial frequency, or a combination of a number of distinct features. Since feature integration plays an important role for visual perception it has been extensively studied. Many psychophysical, electrophysiological, and computational models have explored this complex process from many different angles. In this chapter, a computational model of feature integration, capable of detecting contours defined on the basis of alignment, spatial frequency, or a combination of these two features, is discussed.

First, three psychophysical studies of particular interest to this thesis will be summarized. An overview of contour integration models, alongside a motivation of the model proposed in this chapter will be given. The model is then described, and a set of results exploring contour integration under a large number of different conditions is reported. Finally, a discussion outlining a future direction for this model, and the contributions which have been made to the field, is provided.

4.2 Psychophysical Experiments

Feature integration on the basis of "good continuation", the requirement of elements to follow a smooth global path, has been studied for almost a century by vision scientists (Wertheimer, 1923). As mentioned in section 2.3.1, the linking of elements which follow a smooth global path is one of the basic ways in which we bring order to a visual stimulus. By performing this linking of elements we can distinguish contours and figures. Contour integration builds on the heuristic of good continuation proposed by Gestalt psychologists (Wertheimer, 1923), and adds the requirement of similarity between the local orientation of elements. Literature on contour integration is vast

(see Wagemans et al. (2012a) and Wagemans et al. (2012b) for a review on this topic), as this is a fundamental visual process. Most contour integration paradigms employ gabor patches as edge elements since their properties resemble those which excite cells in the early visual cortex (see section 2.2.1.1). In fact, Gabors have been used as models of simple (Cope et al., 2009) and complex (Spitzer and Hochstein, 1985) cell receptive fields. Given that Gabors have a spatial frequency component, and that spatial frequency is one of the features to which V1 neurons are tuned (see section 2.2.2), it is surprising that not many feature integration studies have looked at the effects of both alignment and spatial frequency similarities/dissimilarities between elements.

Three psychophysical studies (Persike and Meinhardt, 2015a,b; Persike et al., 2009) which combine alignment and spatial frequency in feature integration tasks are of particular interests for this section. The model presented in this section aims to provide a framework of putative neural mechanisms which could be the basis for results observed in these experimental studies. In all variations of the experiments conducted in these three papers, subjects were presented with a two alternative forced choice paradigm (2AFC). In this, observers were presented with two subsequent stimuli and their task was to identify in which of the two stimuli a contour had been presented. In each of the experiments conducted, the contours were generated either by manipulating the spatial frequency component of the elements in the display, the alignment between individual elements, or a combination of these two properties. Following, a coarse description of the experiments, and experimental results, of these three papers will be provided. It is important to keep in mind that this description is intended to inform the reader on the aspects of the experiments which are relevant for the model to be discussed. Thus, some aspects of the papers may be omitted. If an aspect of an experiment, or the results of said experiment, is omitted in this section but becomes relevant later on, it will be mentioned at that later point.

In (Persike et al., 2009) two experiments were conducted. In Experiment One the target stimuli consisted of displays of a field of gabor patches which contained aligned contours of two types, either closed contours in the form of a circle (dubbed O-Shape), or open contours in the form of an S (dubbed S-Shape). The tilt angle of contour elements was varied in order to decrease the visibility of the contours and obtain psychometric curves. Furthermore, the spatial frequency of the Gabors in the display could be varied at three different jitter levels (in combination with the tilt angle of the contour). The levels were UNI containing no spatial frequency jitter of elements; MEDIUM containing a spatial frequency jitter sampled from a uniform distribution spanning a 2 octave interval; and HIGH containing a spatial frequency jitter sampled from a uniform distribution spanning a 2.9 octaves interval (see figure A.1 in the appendix for an example of the target stimuli used in the experiment, or figure 4.1 panel **B**: Jitter on all, for a prototypical example of the stimuli). Six psychometric curves were obtained (see figure A.2). These derived from the combinations of the two shapes (S & O), and the three spatial frequency levels. From these curves, the main result of this experiment can be summarized: observers' ability to correctly report in which stimulus of the 2AFC they had seen the contour decreased with increasing spatial frequency jitters, and/or increasing tilt angle of the contours.

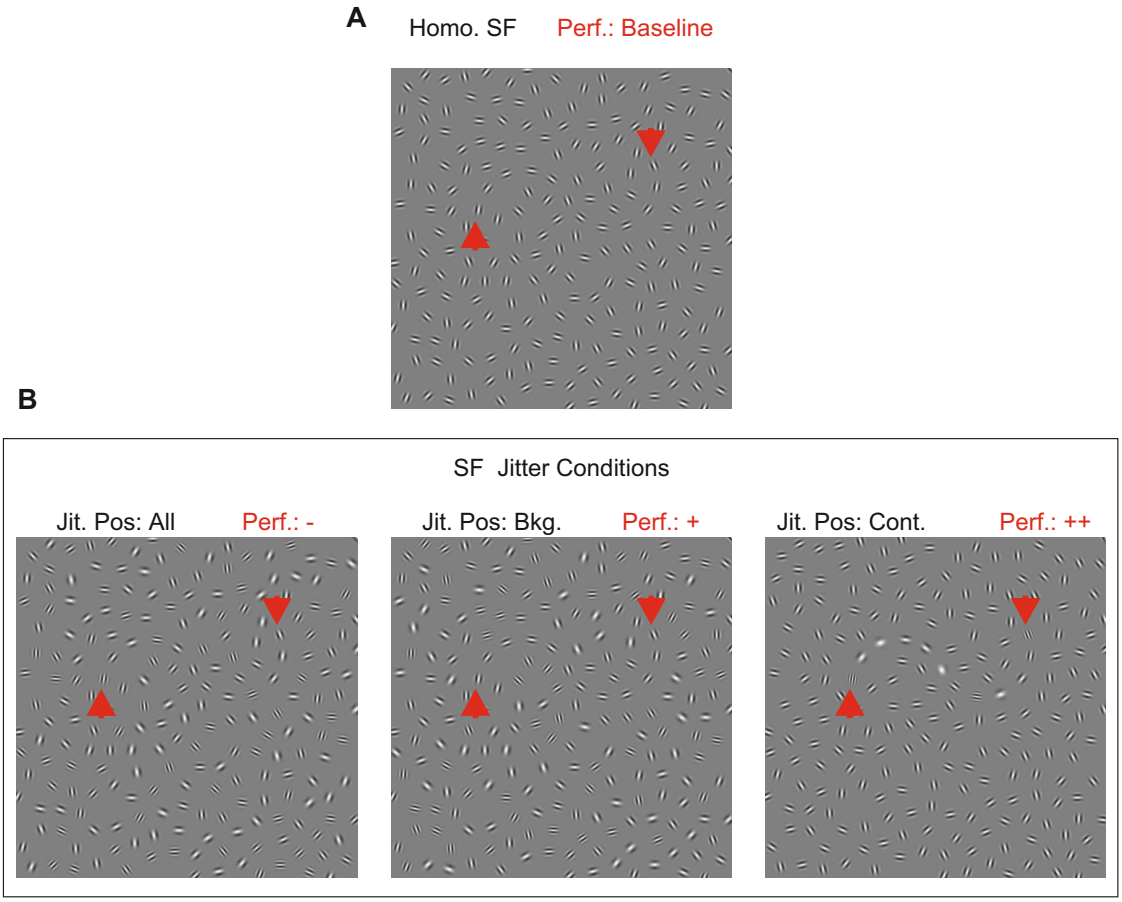

Figure 4.1: Prototypical examples of stimuli used in Persike and Meinhardt (2015b); Persike et al. (2009). Perf.: +/- indicates whether performance increased or decreased in this condition. The use of two plus signs indicates a stronger increase in performance. **A**: Aligned S-Shaped stimulus with a homogeneous SF for all elements. In Persike and Meinhardt (2015b) this type of stimulus was used as a baseline for comparisons of performance increments/decrements. **B**: Aligned contours with spatial frequency jitter applied to the contour, background, or all elements (as indicated for each stimulus). Note the following: **1.** Markers show the position of the contour. These were not shown to subjects in the experiment. **2.** The spatial frequency on contour condition was not performed in any of the experiments being described. This condition was suggested to Persike in private communications. Data for this condition remains unpublished.

In Experiment Two of Persike et al. (2009) contour stimuli could again have an O or S shape. However, either circular elements with a radially symmetric micropattern were used as elements in the display, or gabor patches which were given random orientations along the contour and background (see figure A.3 for a sample of these stimuli). The radial stimuli and the randomly oriented gabor patches were used to ensure that no information regarding the contour could be derived from alignment cues. In this experiment, contours could only be detected based on differences in the spatial frequency properties of the contour elements and the background elements. In all stimulus display types the spatial frequencies of background elements were sampled from a uniform distribution spanning a 2.9 octave range (same as the HIGH condition of Experiment One). To create target stimuli, contour elements were assigned a spatial

frequency which was equivalent to the expected value of the spatial frequency assigned to background elements. In order to reduce detectability of contours, the spatial frequency of contour elements was jittered at different levels (δ). Thus, moving away from a heterogeneous background and homogeneous contour situation, to a situation with a heterogeneous field of elements. The main results of this experiment revealed that observers are generally unable to detect contours based on spatial frequency homogeneity alone (see figure A.4 for a summary of these results).

In Persike and Meinhardt (2015b) very similar experiments to (Persike et al., 2009) were conducted. Only S-shaped contours were used for this experiment, and only gabor patches were used as elements for the stimuli. Performance, in terms of contour detectability, was measured under five distinct spatial frequency jitter conditions (and a number of distinct tilt angles for the contour elements). Spatial frequencies for each individual element in the display were drawn from a uniform distribution, spanning a distinct number of octaves for each experiment. The combinations of spatial frequency jitters used - expressed in octaves - for contour elements (from now on abbreviated as $\boldsymbol{C}$) and background elements (from now on abbreviated as $\boldsymbol{B}$) for each experiment were the following:

(i) $\boldsymbol{C}$: 0.0_{oct} $\boldsymbol{B}$: 0.0_{oct} (iii) $\boldsymbol{C}$: 0.0_{oct} $\boldsymbol{B}$: 3.0_{oct} (v) $\boldsymbol{C}$: 3.0_{oct} $\boldsymbol{B}$: 3.0_{oct}

(ii) $\boldsymbol{C}$: 0.0_{oct} $\boldsymbol{B}$: 2.0_{oct} (iv) $\boldsymbol{C}$: 2.0_{oct} $\boldsymbol{B}$: 2.0_{oct}

(see figure A.5 for a sample of the stimuli used, and figure A.6 for the main experimental results, or figure 4.1 fir a prototypical example of these stimuli). The main results of these experiments confirmed that contours which are defined on the basis of spatial frequencies alone are barely visible. That there might be a signal to noise ratio increase for the case of homogeneous contours in heterogeneous backgrounds (indicated by the steepness of the corresponding psychometric curves), and that contour integration on the basis of alignment is a robust process which is not greatly impaired by spatial frequency dissimilarities.

Finally, in Persike and Meinhardt (2015a), the effects of feature summation (for spatial frequency and orientation) in the process of feature integration was evaluated on aligned contours (O- and S-shaped) generated with gabor elements which could either have a lower or higher spatial frequency than background elements. Two experiments were conducted, in Experiment One, the spatial frequency of contour elements was either shifted up or down in relation to background elements. In Experiment Two, the spatial frequency of the contour remained constant while that of background elements was either shifted up or down (see figure A.7 for a sample of these stimuli, or figure 4.2 for prototypical examples of these stimuli). Unlike in (Persike et al., 2009) and in (Persike et al., 2009) spatial frequency differences between contour elements and background elements were not realized by jittering the spatial frequencies of individual elements, but rather by placing a constant shift on the elements which were being manipulated. In all experiments, four visibility level baselines (i.e.: the feature level at which detection performance was at a desired level) were obtained for contours either defined by: **a.** orientation alone; or **b.** by a spatial frequency difference between contour and background elements alone.

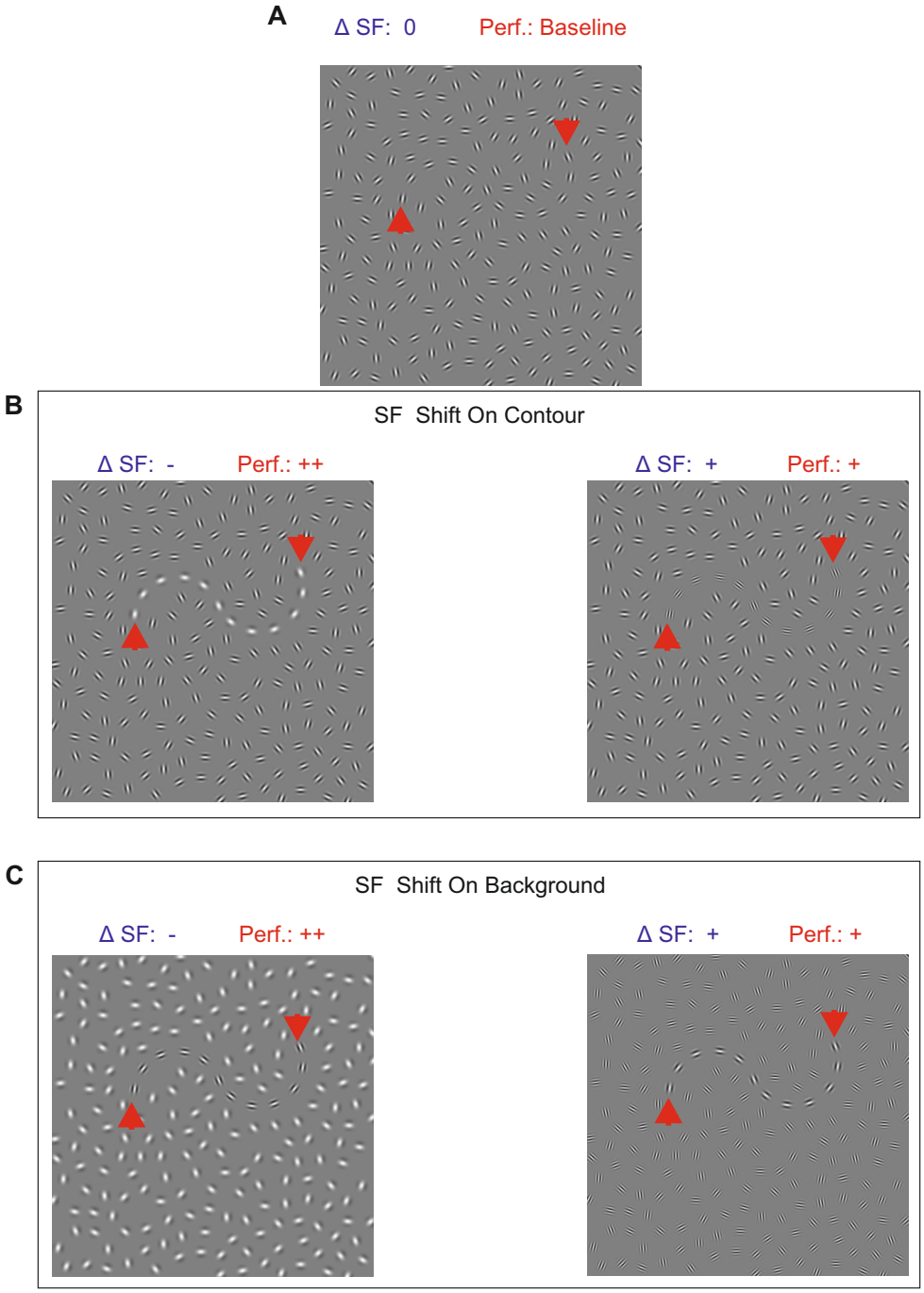

Figure 4.2: Prototypical examples of stimuli used in Persike and Meinhardt (2015a). ΔSF: +/- indicates whether an upwards or downwards spatial frequency shift is shown in the stimulus (in relation to elements in panel **A**). All other conventions are as in figure 4.1. **A**: Aligned S-Shaped stimulus with a homogeneous SF for all elements. **B**: Aligned contours with a spatial frequency shift on contour elements. **C**: Aligned contours with a spatial frequency shift on background elements.

Later on, these two features were combined at each of the visibility levels. Results revealed that improvements in contour detection performance provided by the combination of the two features were greater than linear summation expectations, and were limited by ceiling effects (see figure A.8 for a summary of these results). Furthermore,

these experiments revealed that observers needed less feature contrast for contours with a downwards spatial frequency shift, than vice versa, to reach the same visibility levels.

4.3 Motivation for a Computational Model and Overview of Established Models

A model of feature integration which is sensitive to alignment and spatial frequency (dis)similarities will be presented. The central hypothesis is that feature detectors with similarly preferred (and aligned) orientations, and similarly preferred spatial frequencies, have stronger interactions than those of feature detectors which differ in their preference to these two features.

The model was built in order to study the interactions of these two features in a system which aims to mimic the proposed mechanisms of contour integration (see e.g.: Field et al. (1993); Kovacs (1996)). The psychophysics results outlined in section 4.2 provide an overview of how these two features can affect an observer's performance in detecting contours. However, to the knowledge of the author, up to the point of writing of this thesis, no complete model of contour integration which accounts for the interaction of these two features has been proposed. It is desirable to have such a model in order to further understand the mechanisms underlying visual information processing, and thus be able to generate testable hypotheses on the expected perceptual behaviour of observers when presented with novel stimuli.

Previous models of feature integration have been based on different principles depending on their aims. However, they often share key features. One of the key features which they share is that most models of feature integration aim to create a saliency map in which elements, or areas, in a visual scene are assigned salience values. Typically, the higher this value, the more phenomenologically salient an element (or area) is assumed to be. Through read-out mechanisms, the output of a model, when stimulated with different stimuli, can be compared to determine whether one of the particular stimuli leads to an overall higher salience, or contains elements (or areas) which lead to higher salience than the background. Given that typically, by design, higher salience is rooted in features or areas of interest in a stimulus, read-out mechanisms often judge the stimulus which leads to the higher salience values to be a "target stimulus", and the stimulus which leads to the lower salience values to be a "distractor stimulus".

Another key feature which many feature integration/contour integration models share (as well as many other neural computational models) is that their dynamics are described in terms of population activities. A classical model for neural population dynamics is Wilson and Cowan (1972). In this paper, Wilson and Cowan (1972) described the interaction dynamics of recurrently connected, homogeneous populations of excitatory and inhibitory neurons, in spatially localized areas. Employing just two coupled integro-differential equations, they were able to capture the population dynamics for a large scale neural network. They also showed that their model can be evaluated in a time coarse-grained manner to describe the activity of a large neural network, without losing much temporal accuracy for high frequency responses (Cowan et al., 2016; Wilson and Cowan, 1972). Furthermore, they showed that by varying two key param-

eters, the strengths of connectivity between excitatory and inhibitory model neurons, and the strength of input to the sub-populations, a number of dynamical behaviours (e.g.: oscillations, hysteresis, limit cycle and multiple stable states), which are often observed in the brain, could be reproduced. Because of the simplicity of the model, and because it can be evaluated in a time coarse-grained manner, Wilson-Cowan equations have become a basic building block for many models which employ neural populations, including several contour integration models.

4.3.1 Feature Integration Models

There are different approaches to constructing a feature integration model. One approach, the bottom-up approach, sets out to construct a model based on the known response properties of neurons and on the architecture of the visual system. These type of models are usually also referred to as neural models. A second approach is the top-down approach, in this, a model is constructed to reproduce specific functions of brain regions. Often, while constructing these type of models biological constraints are not of prime importance. Rather than to build a model based on know biological properties, top-down models are built to achieve a goal, and the most efficient mechanisms to achieve said goal are implemented. Once the model is built and it reproduces the phenomenon intended, it might be possible to link the mechanisms to neural substrates, thus giving the model a degree of biological realism. Often, contour integration models constructed with a top-down approach are generative models of a probabilistic nature.

4.3.1.1 Neural Models

One of the most influential models of feature integration constructed with a bottom-up approach is Li (1998)'s contour integration model. (Li, 1998) was one of the first to propose a model of contour enhancement based solely on the known characteristics of V1. Her model is based on V1's physiology, and on V1 cell's tuning properties. It models orientation selective cells, local recurrent neural circuits, and horizontal connections. Visual input is modelled as arriving at discrete spatial locations, and at each spatial location a V1 hypercolumn is modelled. Each hypercolumn is composed of a number of neuron pairs with a receptive field centered at the visuotopic location which stimulates the hypercolumn, and with a preferred orientation. Through a hypercolumn, each of the neuron pairs has a distinct orientation preference which varies in a continuous manner (i.e.: $\theta = k\pi/K$ - where θ is the orientation preference of the neuron pair K, and $k = 1, 2...K$). Each of the neuron pairs consists of an excitatory and an inhibitory neuron that are connected with each other, and each neuron pair is (envisioned as and called) a neural representation of an edge segment. In each of the neuron pairs the excitatory cell receives the visual input and the inhibitory cell is treated as an interneuron. The output of the excitatory cell represents the salience of an edge segment in the input. The cells for any given edge segment are recurrently connected. That is, they can send their outputs to each other, and an edge segment j can excite the modelled neurons of another edge segment i in a monosynaptical manner, by sending their excitatory signal J to the excitatory cell in edge i. They can also inhibit the

edge disynaptically, by directing an excitatory signal W to the inhibitory cell of edge i. Thus, J and W model the synaptic strengths of horizontal cortical connections (see figure 4.3 for a visual representation of the neural connections in a simplified grid in Li (1998)). The main feature of the dynamics in Li (1998) is that they are oscillatory, and that the activities of individual edge segments which are similarly oriented (thus forming a contour) amplify each other and oscillate in synchrony. This is achieved by the assumption that horizontal connections between the neural population representing two edge segments are more likely to target excitatory neurons if their corresponding preferred orientations are aligned. With a similar logic, it is assumed that horizontal connections will target inhibitory cells if their preferred orientations are orthogonal.

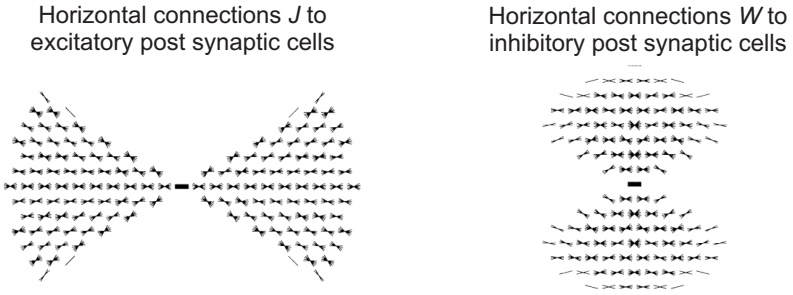

Figure 4.3: Li (1998) model connections. Connections of a horizontal edge (thick line at the center) to all other edges in the visual field with different orientations and positions. Left: horizontal connections from excitatory to excitatory post synaptic cells targeting collinearly oriented edges. Right: horizontal connections from excitatory cells to inhibitory post synaptic cells (of another neuron) targeting parallel oriented edges. Figure obtained from Li (1998).

Others have also constructed contour integration models with a bottom-up approach. Yen and Finkel (1998) proposed a model based on hypercolumns made up of pyramidal cells and interneurons. The cells in a hypercolumn were highly interconnected and no delay was given to the interactions between cells in a hypercolumn. Cells in a hypercolumn were connected to cells with the same orientation preference in neighbouring hypercolumns via long-range horizontal connections. What is unique about the connections in the model proposed by Yen and Finkel (1998), is that not only did their cells have high connection strengths between cells on the same axis but in other hypercolumns with the same orientation preference, but that their cells were also strongly coupled to cells with similar orientations which were in parallel axis (see figure 4.4). These connections were used to help achieve synchrony in the neural activity of cells with similarly preferred orientations in different hypercolumns. The extent of connectivity of contours was judged based on the level of synchronized activity between cells, with high synchrony indicating high levels of connectivity and saliency.

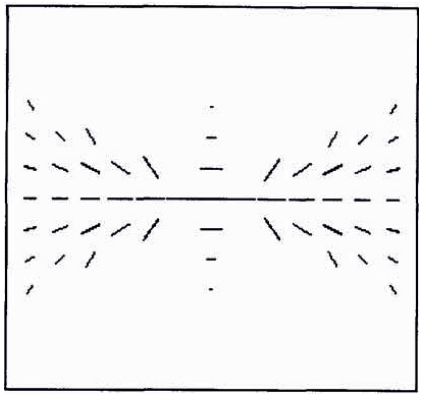

Figure 4.4: Yen and Finkel (1998) model connections. Connectivity pattern of a horizontally oriented cell at the center of the image. The orientation of the lines represents the preferred orientation, while the length of the lines indicates connection strength. Figure obtained from Yen and Finkel (1998).

Pettet et al. (1998) proposed another contour integration model based on a bottom-up approach. They assumed that a model which is stimulated by a field of gabor elements stimulated an equal number of neural units to the number of gabor elements in the field. In their model, each gabor element stimulated one neural unit which was optimally tuned to the orientation of the gabor element. The neural units also received facilitatory input from the other units in the display (with higher inputs from units with a similar orientation preference which were in close proximity to the unit receiving the input). Inhibition was used to help reach convergence in the network, and the parameters of the model were selected so that the model could reproduce psychophysical experimental results.

VanRullen et al. (2001) used a spiking neural network to investigate contour integration. They used spike exchanges between collinearly aligned edge detectors to make neural units corresponding to contour elements fire before units corresponding to background elements. This mechanism is interesting because it posits to explain the rapidity of contour integration.

Ursino and La Cara (2004) focused on V1 and showed that local processing in V1 alone, without influences from higher visual areas, can lead to the augmentation of contour signals. They modelled: **a.** feed forward input from the LGN; **b.** inhibitory feed-forward input to suppress signals originating from neural units which do not share similar orientation preferences; **c.** excitatory cortical feedback; and **d.** a long-range isotropic feedback inhibition used to suppress signals arising from short contours and noise in the network. Even when stimulates for just 30ms, their model was capable of extracting contours from natural and artificial images.

As seen from the results reported above, it is possible to build neuronal models (with a bottom-up approach) which perform contour integration based on several different mechanisms. Thus, it is possible, that the visual system employs several of these mechanisms in order to reliably perform contour integration. Rather than CI being performed by one mechanism, it could be that the visual system performs CI through different mechanisms simultaneously. If this is the case, this would help explain the robustness of CI.

4.3.1.2 Generative Models

Unlike neuronal models, generative models consider perception as an inference process and are often statistical models which specify how a stimulus might be generated from the presence (or absence) of elementary objects in a scene (Ernst et al., 2012). Multisensory cue integration has been modelled in this manner (Ernst and Banks, 2002), and contour integration has also been modelled this way (Ernst et al., 2012). Ernst et al. (2012) is of particular interest to this thesis as it is a precursor to the model to be discussed in this chapter. In order to understand Ernst et al. (2012)'s model, one must first have an overview of Williams and Thornber (2001)'s model, as Ernst et al. (2012) adapted their framework to create their generative model.

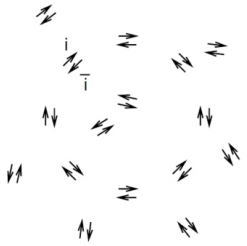

Figure 4.5: Representation of the 2 states (i and $\bar{i}$) in which each edge element can be in the input pattern. The two states correspond to opposite directions in which a particle in random motion can visit an edge element. Figure obtained from Williams and Thornber (2001).

Williams and Thornber (2001) aimed to analyze the properties of illusory contour formation. To do this, they modelled particles moving with constant speeds in directions given by Brownian motion (in a 2 dimensional framework). The particles stochastically visited edge elements, which had a given position and orientation. Each edge element could be in two possible states, denoted as i and $\bar{i}$, which have opposite directions Φ and $\Phi + \pi$ (see figure 4.5). These two states (i and $\bar{i}$) represent the two possible directions from which a particle undergoing a random motion can visit an edge. Knowing the position, and states of all edge elements, the conditional probability that a particle moves from edge i to edge j can be computed. This takes the value of $P(i|j)$[1]. Thus, $P(ij)$ gives the likelihood that two edge elements i and j are connected, and from this the emergence of (illusory) contours can be inferred.

Ernst et al. (2012) extended Williams and Thornber (2001)'s model to create a contour integration model for contours of a finite length N. They defined edge elements e as having a position in a two dimensional plane, and a direction ϕ (hence $e = \{x, y, \phi\}$), and denoted the probability that a contour which passes through e_i will then pass through e_j as $A(e_j|e_i)$. They generated contours of length N by positioning a starting edge at a random position in the association field A, and sampling a sequence of $N-1$ edge elements from the association field A. A was dependent on a radial part

[1]The details the calculations of $P(i|j)$ go beyond the scope of this summary section, for an overview of these calculations please refer to (Williams and Thornber, 2001).

(in which longer radial distances lead to weaker likelihoods of connectivity between elements), and an angular part (in which similar orientations lead to higher likelihoods of connections between elements)[2]. The model generated a probability density function (in essence a link probability between the different elements in a display) which was interpreted as an association field akin to that described by Field et al. (1993). Finally, performance for the model was assessed by presenting it with two stimuli, a target stimulus in which a contour was embedded, and a distractor stimulus in which no contour was embedded. By stimulating the model with these two stimuli, two probability density functions/association fields were obtained. With the aid of a read-out mechanism (see section 4.5.1 for an explanation of read-out mechanisms) applied to the two association fields, a 2AFC task was performed and a performance score was obtained.

Ernst et al. (2012) presented their model with a number of stimuli with different characteristics, and computed a contour detection performance score. A set of parameters which yielded a close matched human performance was found. Thus, Ernst et al. (2012)'s model reproduced human results, provided insight into the possible mechanisms which might be invoked in contour integration, and allowed the authors to make testable predictions on expected human perception of certain stimuli.

Geisler et al. (2001) also took a top-down approach to develop their contour integration model. They studied image statistics in natural images, based on the assumption that these statistics are the source which drove our visual system to generate grouping principles. Thus, they built a model disregarding the biology of the visual system, but based on what they believed drove our visual system to evolve the way it did. In their paper, Geisler et al. (2001) quantified the co-occurrence of edge elements in natural images, and the co-occurrence of edge elements which belong to a particular contour in the image. They showed that they could predict human contour integration performance based on a model which used the statistics which they extracted from natural images to generate a link probability between edge elements. If the link probability between a number of edge elements exceeded a threshold, they deemed the connection of said elements to be a contour.

The model to be presented in the next section takes elements from both, neuronal models and generative models. Furthermore, it was also developed with a blend of top-down and bottom-up approaches. It is presented as a simple neuronal model, as it aims to perform contour integration based on biologically plausible mechanisms.

4.4 The Model

A structurally simple neuronal model based on properties of the early visual system was developed. The purpose of the model was to provide putative explanations for the mechanisms which underlie a number of psychophysical phenomena. In the following subsections, a description of the model will be provided.

[2]For details of the definition of the association field please refer to (Ernst et al., 2012) or section 4.4.1.

4.4.1 General Model Description

Note that, unless otherwise stated, all figures presented in the following sub-sections were generated with the following parameters:

Directionality: Unidirectional	$\lambda_f = 1.45$	$\lambda_{ex} = 0.5$	$\lambda_{in} = 0.6$
$w_0^{ex} = 30$	$w_0^{in} = 1.8$	$\sigma_\alpha = 0.2$	$\sigma_\beta = 0.8$

The model was developed to be stimulated by ensembles E composed of N edge elements. Each edge element, $\vec{e}$, was modelled as consisting of a location $\vec{r}$, an orientation ϕ, and a spatial frequency F, so that $\vec{e}_i = \{\vec{r}_i, \phi_i, F_i\}$ and $E = \{\vec{e}_i\}$ for $i = 1...N$. In each ensemble E, N was composed of background elements, N_b, and contour elements, N_c, so that $N = N_b + N_c$. In target ensembles $N_c = 12$ and in distractor ensembles $N_c = 0$, on average $N = 225$. See figure 4.6 for a sample of a distractor and different types of target stimuli.

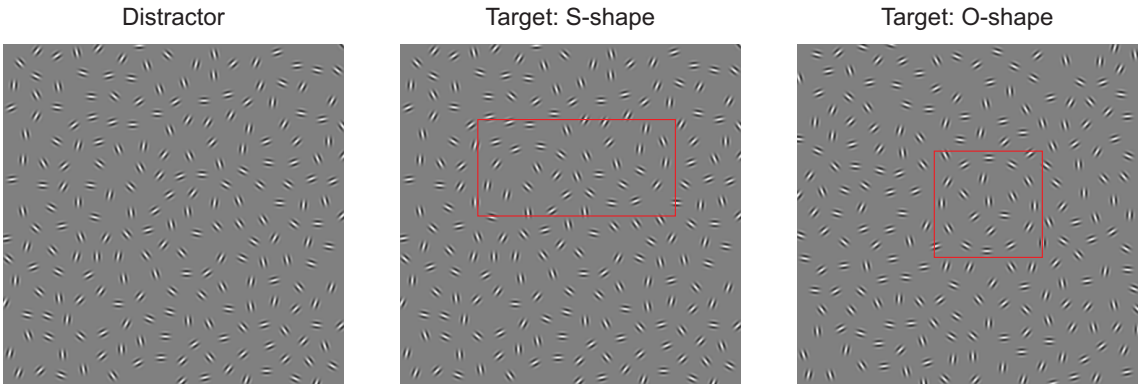

Figure 4.6: Examples of the stimuli with which the model was stimulated. The red rectangles in the two target trials were added to the stimuli to help the reader find the contour. They were not part of the stimuli used in the psychophysical experiment, nor of the stimuli used to stimulate the model.

Each edge element in the ensemble was represented by the activation of one or two neural populations P (see section 4.4.2 for a detail explanation of the two cases), with a receptive field centered at $\vec{r}_i = \{x_i, y_i\}$, a preferred orientation ϕ_i, and a preferred spatial frequency F_i. Neural populations with RFs properties which did not match the stimulus presented to the model were assumed to not be activated. Thus, these neural populations were not modelled.

The activation dynamics are described by time coarse-grained Wilson-Cowan equations (Wilson and Cowan, 1972):

$$\tau \frac{dA_i(t)}{dt} = -A_i(t) + g[J_i^{rec}(t) + J_i^{ext}(t) + \eta_i(t)] \tag{4.1}$$

Each modelled neuronal population has an activity of $A_i(t)$; and $g(x)$ is a rectification function: $g(x) = 0$ for $x < 0$ and $g(x) = x$ for $x \geq 0$. Without loss of generality τ can be set to 1 and time can be re-scaled in order to obtain all possible solutions of

$A_i(t)$, for any given τ. The feedforward input, J_i^{ext}, is constant and can thus be set to 1 w.l.o.g.

$\eta_i(t)$ denotes uncorrelated Gaussian white noise realized as:

$$\eta_i = \sigma^{Noise}\xi\sqrt{T} \tag{4.2}$$

where σ^{Noise} is a constant, ξ is a normal distributed random variable with mean of 0 and a standard deviation of 1, and T is the integration time-step used in the numerical solution of eq. 4.1 (set to $0.025ms$). J_i^{rec}, the recurrent input, provides feedback via a coupling matrix $W(E) = \{w_{ik}\}$:

$$J_i^{rec}(t) = \sum_k w_{ik}A_k(t) \tag{4.3}$$

The coupling matrices used to define the recurrent input depend on the particular stimulus configuration E, and are defined as:

$$W(E) = (w_0^{ex}W^{ex}(E) - w_0^{in}W^{inh}(E)) \circ W^f(E) \tag{4.4}$$

For a visual depiction of $W(E)$ please refer to the lower row of figure 4.10.

W^{ex} describes excitatory interactions and W^{inh} describes unspecific inhibitory interactions. W^f describes the modulating interactions given by spatial frequency differences between any two $\vec{e}$ in E. w_0^{ex} and w_0^{in} are constants used to establish a desired ratio between excitatory and inhibitory interactions. $\circ$ indicates an elementwise multiplication.

$W^{ex} = \{w_{ij}^{ex}\}$ consists of: (a) a radial part depending on the distance r_{ij} between $\vec{e}_i$ and $\vec{e}_j$; and (b) an angular part which is dependent on the orientations of $\vec{e}_i$ and $\vec{e}_j$, as well as on the relative positions of $\vec{e}_i$ and $\vec{e}_j$ (quantified in the variables α_{ij} and β_{ij} respectively). For a visual depiction of the different components of W^{ex} please refer to figure 4.7, and for a combination of the components please refer to the upper row of figure 4.10. The architecture of W^{ex} is based on the model proposed by Ernst et al. (2012).

$$W^{ex}(r, \alpha, \beta) := W^{ex,rad}(r) \circ W^{ex,ang}(\alpha, \beta) \tag{4.5}$$

The radial part, $W^{ex,rad}$, of W^{ex} in equation 4.4 is modelled as an alpha function that is zero for self-couplings and has its maximum at $r_{ij} = \lambda_{ex_{ij}}$. Elements of $W^{ex,rad}$ are defined as:

$$w_{ij}^{ex,rad}(r_{ij}) := \frac{1}{4\pi\lambda_{ex_{ij}}^3} r_{ij} \exp\left(-\frac{r_{ij}}{\lambda_{ex_{ij}}}\right) \tag{4.6}$$

where r_{ij} is the Euclidean distance between $\vec{e}_i$ and $\vec{e}_j$, and the pre-factor $\frac{1}{4\pi\lambda_{ex_{ij}}^3}$ is a normalization factor. See figure 4.7 for a visualization of $W^{ex,rad}$.

Two models which differ on their definition of $\lambda_{ex_{ij}}$ will be discussed. These are dubbed the ***Fixed Scaling Model*** and the ***Variable Scaling Model***. In the *Fixed Scaling Model* $\lambda_{ex_{ij}} = \lambda_{ex_0}$, a constant, whereas in the *Variable Scaling Model* $\lambda_{ex_{ij}}$ is

defined as:

$$\lambda_{ex_{ij}} := \lambda_{ex_0} 2^{-\frac{\Delta_{s_i}}{2}} \tag{4.7}$$

where $\Delta_{s_i} = \log_2(F_i) - \log_2(F_0)$. Thus, in the *Variable Scaling Model* $\lambda_{ex_{ij}}$ is defined relative to the length scale λ_{ex_0} of a neuronal population with a preferred spatial frequency of F_0. A variable range for lateral interactions which depends on differential SF preferences of neural units is supported by experimental evidence, as shown by Ernst et al. (2016); Polat and Sagi (1993); Tolhurst and Barfield (1978). The interaction length scale was assumed to depend on the SF preference of the post synaptic neuron. Experimental evidence supports this decision (Polat and Sagi, 1993).

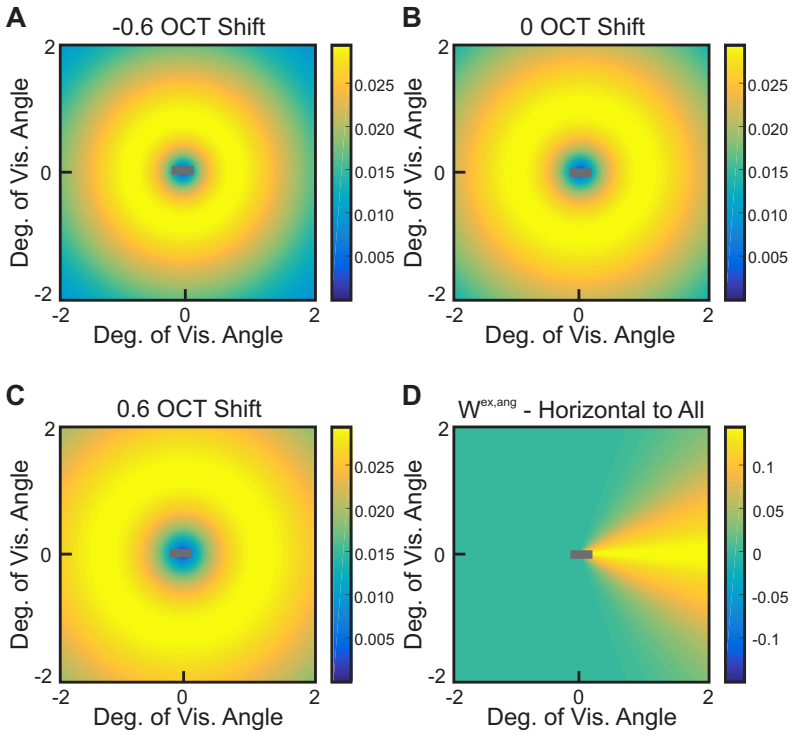

Figure 4.7: **A, B, C**: Connection strengths given by the radial part (eq. 4.6) of the excitatory weights (eq. 4.5). The different effects of spatial frequency shifts in the *Variable Scaling Model* are shown in panel **A** for a negative SF shift, and in panel **C** for a positive SF shift. SF shifts are expressed with respect to f_0 ($f_0 = \log_2(F_0)$). Panel **B** shows the radial part of the association field in a 0 octave shift case (i.e.: the SF preference of all neural units was F_0) in the *Variable Scaling Model*, or the radial part in all cases in the *Fixed Scaling Model*. **D**: Angular part (eq. 4.10) of the excitatory weights. The strength of connections of a horizontal element to all other orientations is shown for a unidirectional association field (see section 4.4.2 for a comparison of a unidirectional and bidirectional association field). All plots use an edge element (shown by the grey bar in the center of the display), which stimulates a neural population with an RF centered at that visuotopic location, as a presynaptic origin.

The convention for a presynaptic origin will be used in all further plots in which a

grey bar is presented at the center of the plot.

As all aspects of the *Variable Scaling Model* and the *Fixed Scaling Model* are identical, with the exception of the definition of $\lambda_{ex_{ij}}$, a common description of all other elements of the model continues.

In equation 4.4 elements of W^{inh} are modelled as:

$$w_{ij}^{inh} = \frac{1}{4\pi\lambda_{in}^3} r_{ij} \exp\left(-\frac{r_{ij}}{\lambda_{in}}\right) \tag{4.8}$$

and elements w_{ij}^f of matrix W^f are modelled as $w_{ij}^f = f_{exp}(F_i, F_j)$ with

$$f_{exp}(F_i, F_j) = \exp\left(-\frac{|\log_2(F_i) - \log_2(F_j)|}{\lambda_f}\right) \tag{4.9}$$

where λ_{in} and λ_f are constants. See figure 4.8 for a visualization of W^{inh} and W^f.

A decaying exponential was chosen to model the connection strengths given by SF (dis)similarities in the preferences of SF by the distinct neural populations. It was assumed that interactions should be strongest when neural populations have the same preferred spatial frequencies.

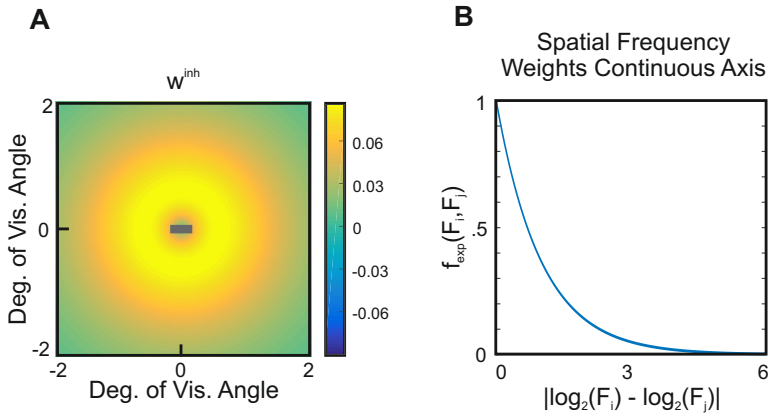

Figure 4.8: **A**: Inhibitory weights (see equation 4.8) in 4.4. Note that the weight values are depicted as positive in the image, however, these values are subtracted in eq. 4.4, thus making them inhibitory. **B**: Depiction of W^f on a continuous axis.

The angular part of W^{ex} is parametrized as

$$W^{ex,ang}(\alpha,\beta) := 0.5 \cdot \left[M\left(\frac{\beta}{2}, \alpha, \frac{1}{\sigma_\alpha^2}\right) M\left(\frac{\beta}{2}, 0, \frac{4}{\sigma_\beta^2}\right) + M\left(\frac{\beta}{2}, \alpha+\pi, \frac{1}{\sigma_\alpha^2}\right) M\left(\frac{\beta}{2}, \pi, \frac{4}{\sigma_\beta^2}\right)\right] \tag{4.10}$$

with σ_α, σ_β being constants, and M a von-Mises function that corresponds to Gaussian distributions defined on a circular support. The parametrization choice of $W^{ex,ang}$ allows for: **a.** the magnitude of excitatory interactions between neuronal populations to decrease on a length scale of σ_α with the distance from a co-circular edge configuration;

b. and for the magnitude of excitatory interactions between neuronal populations to decrease on a length scale σ_β, with an increasing curvature K. Note that for co-circular edge configurations with edge distances r, K is related to β via $K(r) = 2\sin(\beta/2)/r$, and α is related to β via with $2\alpha = \beta$ (Ernst et al., 2012). See figure 4.9 for a visualization of the geometrical relation between elements $\vec{e}_i$ and $\vec{e}_j$ established in equation 4.10.

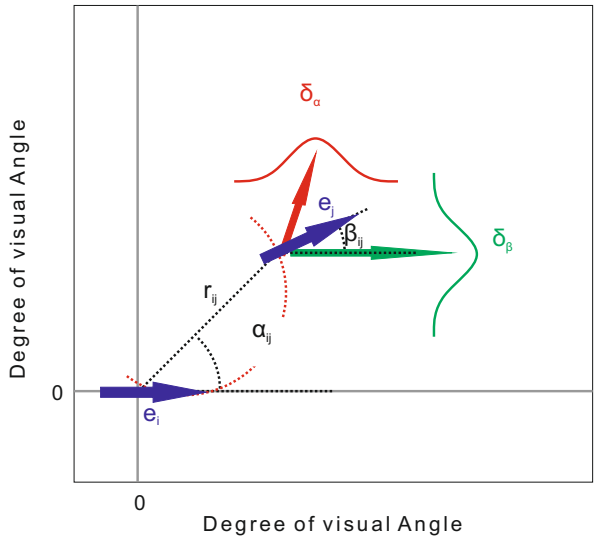

Figure 4.9: Geometric relation between elements $\vec{e}_i$ and $\vec{e}_j = \{x_j, y_j, \phi_j\}$ and their relative coordinates r_{ij}, α_{ij}, β_{ij}. Red arrow: direction in which e_j should be oriented to have a perfect co-circular continuation of a contour through $\vec{e}_i$ and $\vec{e}_j$. The connection strength for the two neural units ij depends on the deviation of ϕ_j from the orientation shown by the red arrow, on the length scale σ_α. Connection strength also depends on the difference β_{ij} between the directions of elements i and j, on a length scale σ_β (Ernst et al., 2012).

M is defined as follows:

$$M(x, \mu, \kappa) = \frac{1}{2\pi I_0(\kappa)} \exp(\kappa cos(x - \mu)) \tag{4.11}$$

$\mu \in [-\pi, \pi]$ is the circular mean, $\kappa > 0$ is a concentration parameter which modulates the reach of all neural populations' excitatory interactions, and $x \in [-\pi, \pi]$ is the angular variable. I_0 is the Bessel function of the first kind, of order 0. By the relationship $\sigma = \sqrt{1/k}$, k is related to the width (σ) of a Gaussian distribution. α and β are given by:

$$\alpha(\vec{e}_i, \vec{e}_j) := \text{atan2}(y_j - y_i, x_j - x_i) - \varphi_i \tag{4.12}$$

$$\beta(\vec{e}_i, \vec{e}_j) := \varphi_j - \varphi_i \tag{4.13}$$

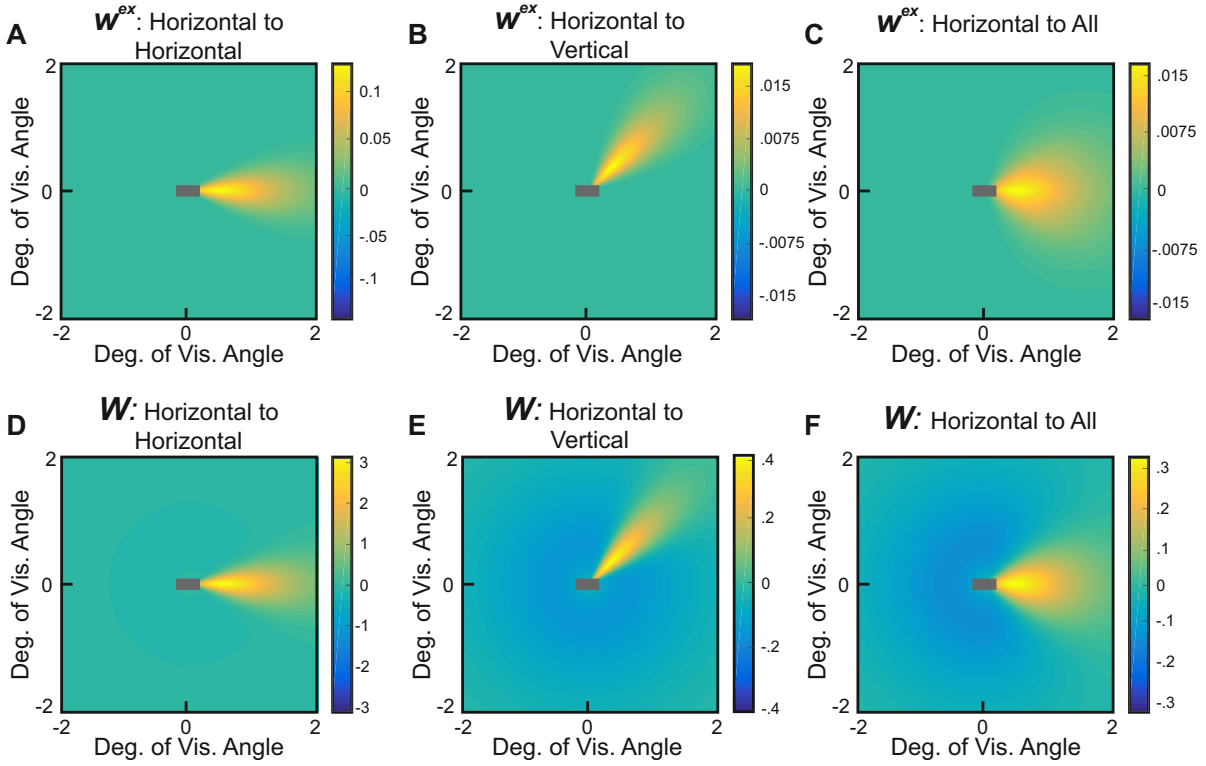

Figure 4.10: Panels **A, B** and **C**: Connection strengths given by W^{ex}, which is composed of the angular (eq. 4.10) and radial (eq. 4.7) parts of the excitatory weights (eq. 4.5) in W (eq. 4.4). Panels **D, E** and **F**: Connection strengths given by W (eq. 4.4). **A** and **D**: Horizontal to horizontal connections; **B** and **E**: Horizontal to vertical connections; **C** and **F**: Horizontal to all orientations connections. The faint blue shade seen in the lower three panels is given by the inhibitory weights (eq. 4.8). The overall connection strengths are higher in panels **D**, **E**, and **F** than in **A**, **B**, and **C**, respectively, because of the influence of W_0^{ex} and W_0^{inh} (see eq. 4.4).

Note that, henceforth, when reference is made to spatial frequency values, the units used will be octaves. Thus, henceforth, f is defined as follows:

$$f = \log_2(F) \tag{4.14}$$

as a consequence, f_0 is defined as $f_0 = \log_2(F_0)$.

4.4.2 Unidirectional and Bidirectional Couplings

Two types of models which differ in the definition of their association fields, by the directionality of the couplings were conceived (see figure 4.11). The first, a model with a unilateral directionality preferences was conceived by modeling two neural populations, P_{i_a} and P_{i_b}, for each line segment in E. Each of the two neural population was modelled with a receptive field centered at the same location $\vec{l_i}$, in a Cartesian plane, however, the two neural populations were assigned axons which are assumed to project in opposite preferred directions. If the preferred projection direction of P_{i_a} was ϕ, then

the preferred projection direction of P_{i_b} was $\phi+\pi$. Thus, for the models with unilateral directionality the number of neural populations modelled, N_P, was $N_P = 2N$ (in order to have neural units which represent all possible projection directional preferences).

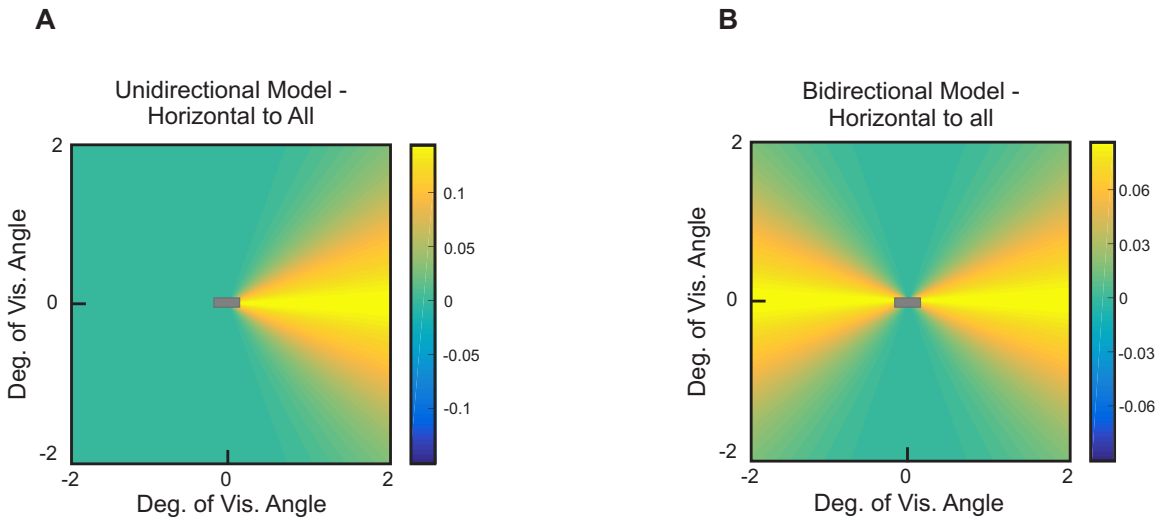

Figure 4.11: Angular part of the association field. Panel **A**: unidirectional model. Panel **B**: bidirectional model. Note that both model images were created with the parameters listed in section 4.4 (except for directionality) in order to have a comparison which only differed in directionality.

The second was a model with a bilateral projection directionality for each of the neural populations. In this, each line segment in E was modelled by just one neural population P_i, with a receptive field center located at $\vec{l_i}$. The coupling structure for each P_i extended bidirectionally from a source population, so that each source population had two preferred projection directions, ϕ_i and $\phi_i + \pi$. In order to obtain the angular part of the couplings with the symmetry just described $\underline{\vec{e}}_i := \{\vec{r}_i, \phi_i + \pi, f_i\}$ was introduced, and $W^{ex,ang,bi}$ was defined as follows:

$$\begin{aligned} W^{ex,ang,bi}(\vec{e}_i, \vec{e}_i) &:= W^{ex,ang}(\vec{e}_i, \vec{e}_i) + W^{ex,ang}(\underline{\vec{e}}_i, \vec{e}_i) \\ &+ W^{ex,ang}(\vec{e}_i, \underline{\vec{e}}_i) + W^{ex,ang}(\underline{\vec{e}}_i, \underline{\vec{e}}_i) \end{aligned} \tag{4.15}$$

The unidirectional model has the advantage that it will not enhance the activity (to a great extent) of neural populations which receive feedforward input from a short number of aligned edge elements (e.g.: 2), if the following element in a chain has a drastic change in orientation. As such, a unidirectional model will have a lower false positive rate than a bidirectional model. However, a unidirectional model is not as biophysically plausible as a bidirectional one. In the brain, contour integration is likely performed by recurrent activation of synaptically linked orientation selective cells with similarly preferred orientations. Horizontal long-ranging axons within the primary visual cortex, as those described by Bosking et al. (1997); Stettler et al. (2002), and/or feedback connections from V2 to V1 (Shmuel et al., 2005; Stettler et al., 2002),

are likely responsible for the linking of individual edge elements into the percept of a contour. These cells have symmetric dendritic trees. Thus, the neural substrates of contour perception are likely akin to the structure of the bidirectional model, and not to the unidirectional model. Biophysically, the unidirectional model would require two distinct neural populations with similar preferred orientations, and receptive field centers located at the same visuotopic location, but with asymmetric dendritic trees (Ernst et al., 2012). This structure is unlikely to exist in the human brain.

4.4.3 Characterization of W's Components

From now on, for simplicity purposes, when a reference is made to neural units stimulated by contour elements, or neural units stimulated by background elements, these will be called contour units, or background units, respectively. Unless otherwise stated, all figures in this section were generated with the following parameters:

$$\begin{array}{llllll} \text{Directionality: Bidirectional} & \lambda_f = 1.15 & \lambda_{ex} = 0.5 & \lambda_{in} = 0.6 \\ w_0^{ex} = 10 & w_0^{in} = 1 & \sigma_\alpha = 0.1 & \sigma_\beta = 0.8 \end{array}$$

In order to understand the response of the model to different stimulus properties (e.g.: alignment of contour elements, SF manipulations), it is helpful to characterize the response of the different components to these properties.

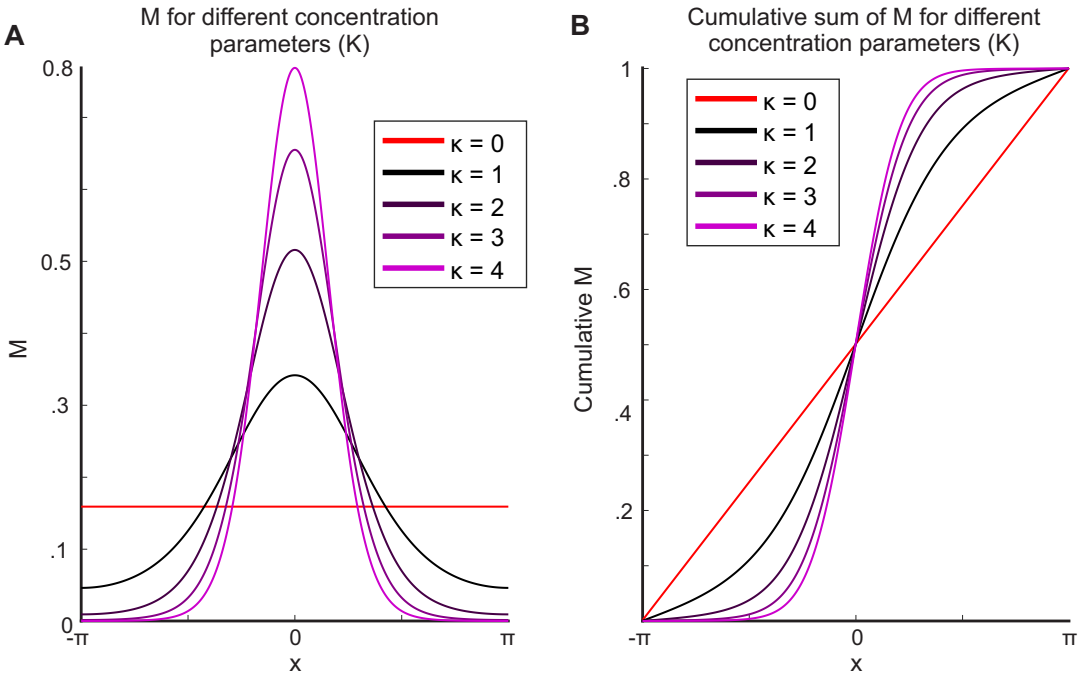

Figure 4.12: **A**: von-Mises probability density functions with a support between $-\pi$ and π, and a mean $\mu = 0$ (see eq. 4.11). **B**: Cumulative sum of functions shown in panel **A**.

The angular part of W^{ex} is based on a von-Mises function (eq. 4.11). As show in figure 4.12 the amplitude of the function increases as the concentration parameter

κ increases. Also, as previously stated, the width (σ) of the function is related to κ ($\sigma = \sqrt{1/\kappa}$).

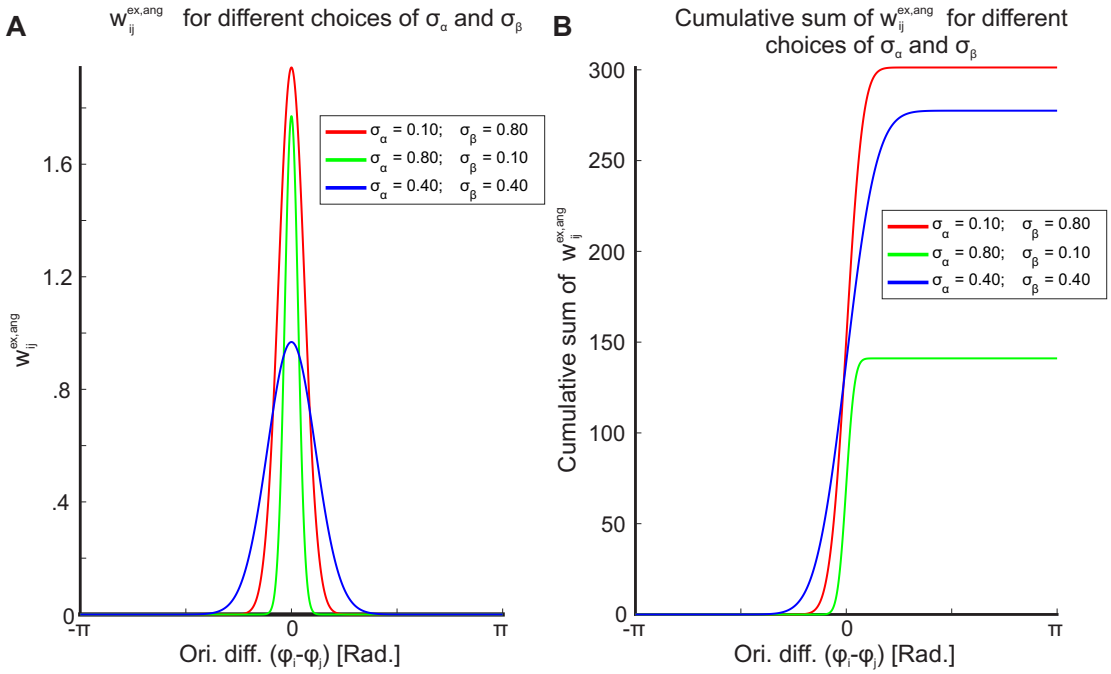

Figure 4.13: **A**: $w_{ij}^{ex,ang}$ for different choices of σ_α and σ_β. **B**: Cumulative sum of functions shown in panel **A**.

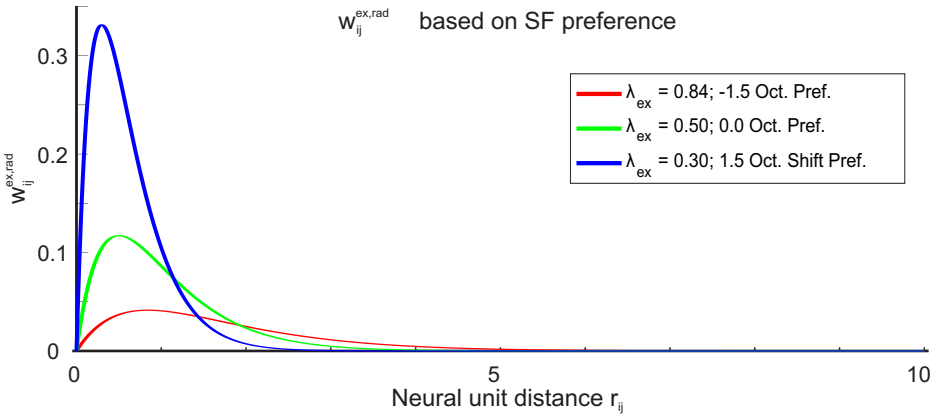

Figure 4.14: $w_{ij}^{ex,rad}$ for different choices of λ_{ex}. The choices of λ_{ex} can be related to the spatial frequency preference of neural units in the the *Variable Scaling Model*. As indicated in the legend, $\lambda_{ex} = 0.84$ is related to an SF preference of -1.5 oct. in relation to f_0; $\lambda_{ex} = 0.5$ is related to an SF preference of f_0; and $\lambda_{ex} = 0.3$ is related to an SF preference of 1.5 oct. in relation to f_0.

Figure 4.13 shows how the amplitude and width of $w_{ij}^{ex,ang}$ changes with different choices for σ_α and σ_β (with the parametrization chosen - eq. 4.10). A choice of $\sigma_\alpha = 0.1$ and $\sigma_\beta = 0.8$ (show in red) allows for larger angle differences to yield high values of $w_{ij}^{ex,ang}$ than the reverse choice (shown in green). This choice of parameters also leads

to higher values of $w_{ij}^{ex,ang}$ for elements with the same orientation, since $w_{ij}^{ex,ang}$ is not normalized (as seen in panel **B**).

Figure 4.14 shows the changes in amplitude and width of $w_{ij}^{ex,rad}$ based on different λ_{ex}. Larger values in this spatial scaling lead to lower amplitudes of $w_{ij}^{ex,rad}$. $w_{ij}^{ex,rad}$s with lower amplitudes have a further reach, implying a further reach of lateral connections. In the *Variable Scaling Model*, shifts to lower spatial frequency preferences lead to lower amplitudes, and vice versa. Thus, the spatial frequency preference of neural units is related to said unit's lateral connection reach.

W^{ex} is composed of $W^{ex,ang}$ and $W^{ex,rad}$, in figure 4.15 W^{ex} is shown for different choices of σ_α and σ_β, and λ_{ex}. As seen in figure 4.15, W^{ex} is most effective for small angle differences and for neural units in close proximity.

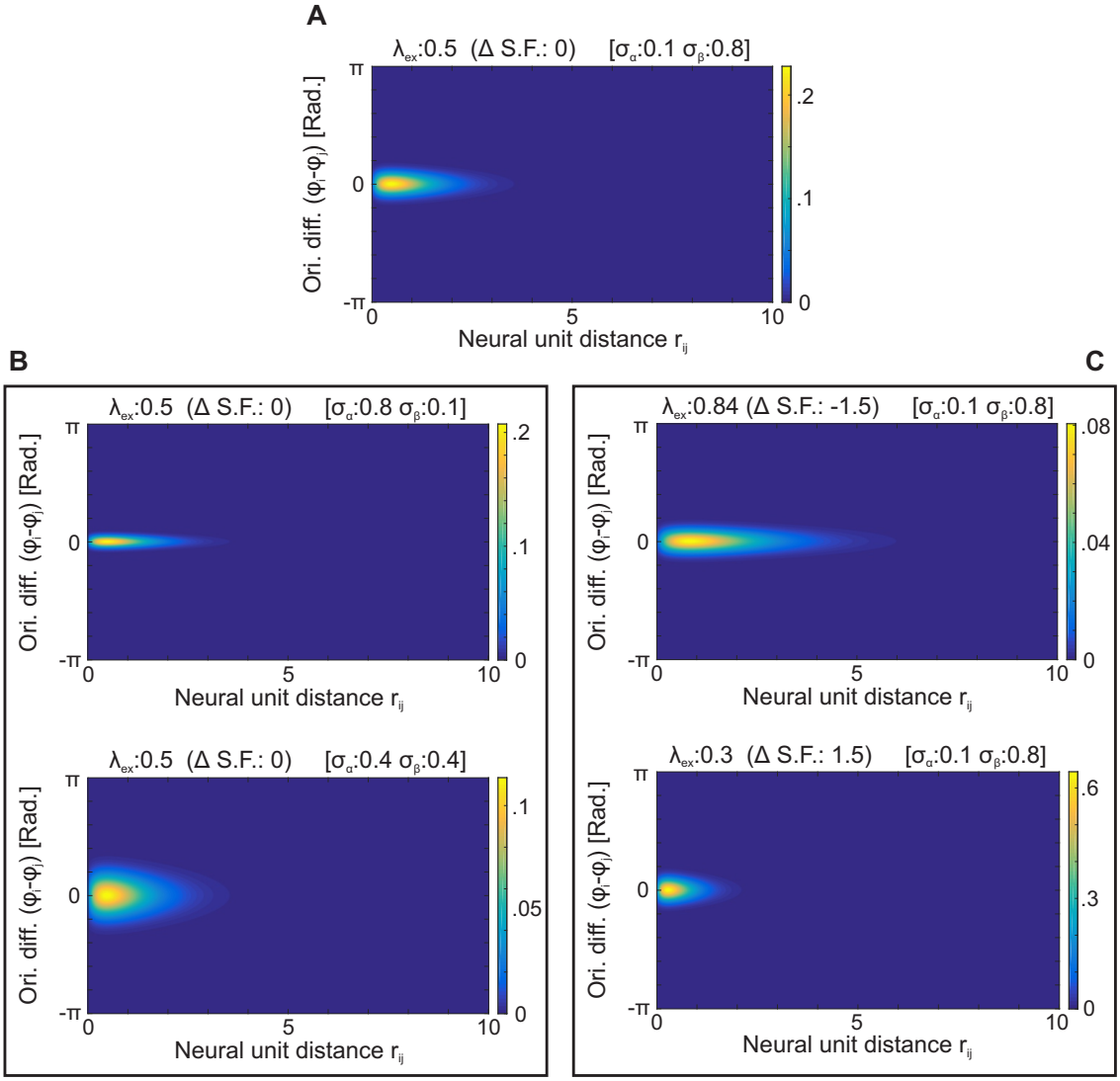

Figure 4.15: w_{ij}^{ex} for different choices of σ_α, σ_β, and λ_{ex}. **A**: $\sigma_\alpha = 0.1$, $\sigma_\beta = 0.8$, and $\lambda_{ex} = 0.5$; **B**: Different choices of σ_α and σ_β, with $\lambda_{ex} = 0.5$. **C**: Different choices of λ_{ex}, with $\sigma_\alpha = 0.1$ and $\sigma_\beta = 0.8$.

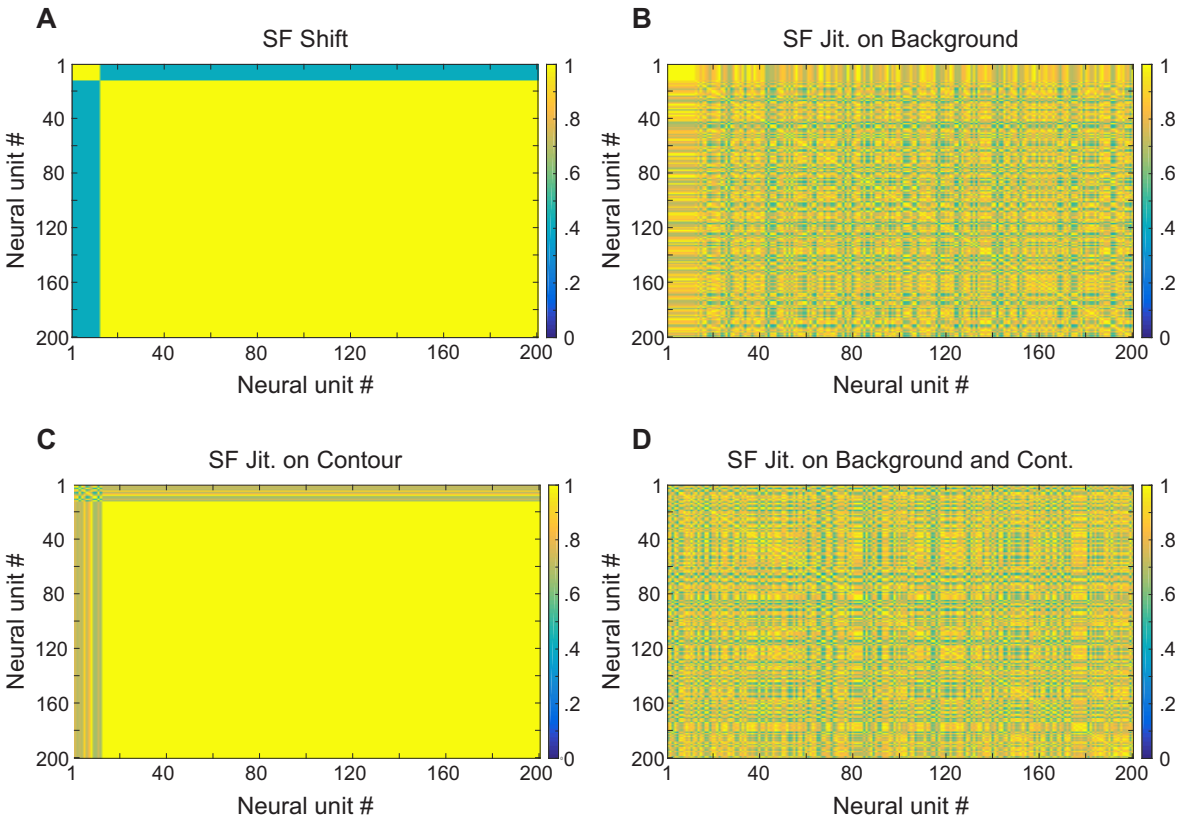

Figure 4.16: Coupling matrices W^f. Note the following: **1.** Neural units stimulated by contour elements are indexed from 1 to 12 (shown in the top left corner of each panel). **2.** The diagonal represents self couplings which were not realized in the model. **A**: 1 oct. difference in SF preference of contour units and background units. **B**: Homogeneous SF preference for contour units, heterogeneous SF preference for background units (SF preference sampled from a uniform distribution with a 1 oct. width for background units). **C**: Heterogeneous SF preference for contour units (SF preference sampled from a uniform distribution with a 1 oct. width for contour units), homogeneous SF preference for background units. **D**: Heterogeneous SF preference for all neural units (SF preference sampled from a uniform distribution with a 1 oct. width).

W^{inh} is the same function as $W^{ex,rad}$, only with a different decay factor (see eq. 4.8 and eq. 4.7 respectively). As seen in $W(E) = (w_0^{ex}W^{ex}(E) - w_0^{in}W^{inh}(E)) \circ W^f(E)$ (eq. 4.4 - the equation used to to define the recurrent input to the model), the factor $(w_0^{ex}W^{ex}(E) - w_0^{in}W^{inh}(E))$ is modulated by $W^f(E)$. The shape of W^f is shown in figure 4.8 panel **B**, and coupling matrices for $W^f(E)$ with different E are shown in figure 4.16.

For W^f, it did not matter whether neural units stimulated by contour or background elements had a higher (or lower) SF preference. As long as there is a (homogeneous) difference in the SF preference between these two groups of neural units, the coupling matrices have the structure shown in figure 4.16, panel **A**. In this case, connections between contour neural units are at a maximum, and connections between background neural units are also at a maximum. However, contour-to-background connections/background-to-contour connections are at the minimum of $W^f(E)$. This

is due to the use of absolute values in W^f (see eq. 4.9). Only the magnitude of the difference in SF preference between neural units i and j is of relevance for w^f_{ij}. That is, it does not matter whether neural unit i has a higher SF preference than neural unit j, or vice versa.

When the SF preferences of either contour or background neural units was sampled from a given distribution, a pattern also emerged. In panel **B**, $W^f(E)$ is shown for a case in which contour units had a homogeneous SF preference, and the SF preference of background units was sampled from a (1 oct. wide)[3] normal distribution. In this case, the connection strengths between contour units are at the maximum of $W^f(E)$, whereas connections between background units range between the maximum and minimum of $W^f(E_B)$ (where E_B is the subset of background neural units in E). Contour-to-background and background-to-contour connection strengths ranged from the maximum and minimum of $W^f(E_{CtoB})$ (where E_{CtoB} is the subset of contour-to-background neural unit connections in W^f). Furthermore, the connections strengths of contour-to-background, and background-to-contour neural units were equal, as $w^f_{ij} = w^f_{ji}$.

In panel **C** the opposite pattern as described above is present. In this case the SF preference of background units was homogeneous, and that of contour units was sampled from a (1 oct. wide) normal distribution. Finally, when the SF preference for all neural units is sampled from a (1 oct. wide) normal distribution, no pattern emerges in $W^f(E)$, other than $w^f_{ij} = w^f_{ji}$. This case is shown in panel **D**.

4.4.4 Characterization of W

As in the previous section, unless otherwise stated, all figures in this section were generated with the following parameters:

Directionality: Bidirectional	$\lambda_f = 1.15$	$\lambda_{ex} = 0.5$	$\lambda_{in} = 0.6$
$w_0^{ex} = 10$	$w_0^{in} = 1$	$\sigma_\alpha = 0.1$	$\sigma_\beta = 0.8$

Visual depictions of $W(E)$, the coupling matrix given by equation 4.4, are shown in figures 4.17 and 4.18 for different E. Figure 4.17 shows cases related to the spatial frequency shift condition, and figure 4.17 shows cases related to the SF jitter condition (see section 4.2 and section 4.6 for details of these conditions). In order to facilitate the discussion of the numerical simulations carried out for the different experiments presented in sections 4.6.1, a discussion of the aforementioned manipulations on E ensues.

Figure 4.17 panel **A** shows that when the model is presented with oriented contours, (on average) W has excitatory connections between contour neural units, and either inhibitory connections, or connection strengths close to 0, for all other connections.

[3]The width of the normal distribution is specified in brackets because it is not of out-most relevance. What is of relevance is that the distribution chosen was a normal distribution with a none zero standard deviation.

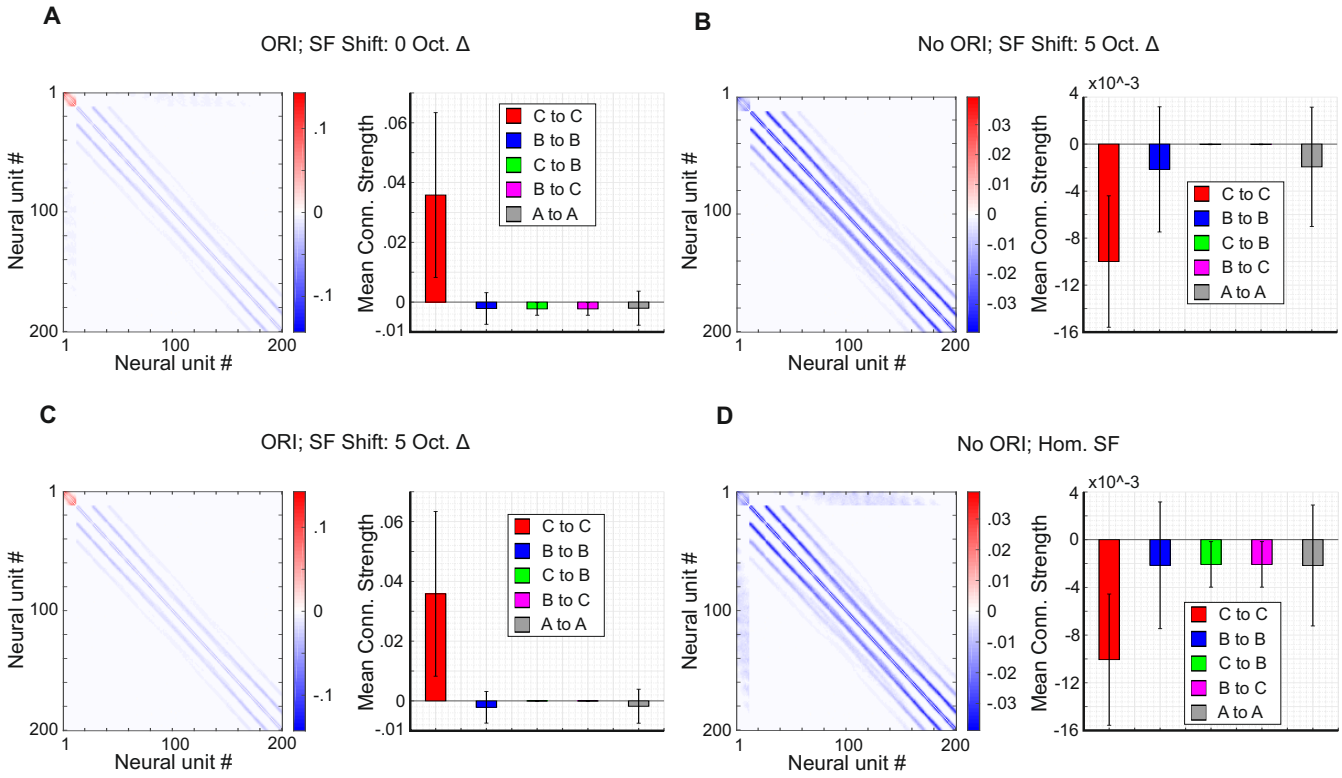

Figure 4.17: Visualization of W and mean connection strengths for different clusters of neural units (as indicated in the plots). In the figures showing mean connection strengths, C to C refers to contour-to-contour connections, B to B refers to background-to-background connections, C to B refers to contour to background connections, B to C refers to background to contour connections, and A to A refers to all-to-all connections. Data represents the average of 1000 samples. Neural units stimulated by contour elements are indexed from 1 to 12 (shown in the top left corner of each panel). Even in cases when no contour was present, the indices 1 to 12 represent would be contour neural units (i.e.: neural units with an RF at the location where a contour would appear if it was defined in E). Δ SF is expressed in relation to f_0. Data was obtained with the *Fixed Scaling Model*. **A**: Oriented contour and a homogeneous SF preference for all neural units. **B**: Random orientation for all elements and an SF preference difference of 5 oct. between contour and background neural units. **C**: Oriented contour and an SF preference difference of 5 oct. between contour and background neural units. **D**: Random orientation for all elements and a homogeneous SF preference for all neural units.

In panel **A**, excitatory connections are seen between contour units because the contour is aligned. Thus, for contour-to-contour connections, $w_{ij}^{ex,ang}$ approaches its maximum (see figure 4.13), and $w_{ij}^{ex,rad}$ (see figure 4.14) is relatively large, since contour neural units have RFs that are in close proximity to each other. Connections between neural units stimulated by nearby elements, which are not contour elements, are mainly inhibitory (these connections are shown in the diagonal of the left figure in all panels). Connection strengths between neural units stimulated by elements which are not in close proximity tend to be close to 0. The strength of these connections approaches 0 because: **a.** $w_{ij}^{ex} \to 0$ due to the influence of the radial part of w_{ij}^{ex} (see eq. 4.6 and

figure 4.14); and **b.** inhibitory connections also decay with distance (see eq. 4.8), thus w_{ij}^{inh} approaches 0 with large distances.

Panel **B** (of figure 4.17) shows that when there is no oriented contour, **and** the SF preference of contour units and background units differ, connections are (on average) inhibitory. This is because in W, excitatory connections are only given by $W^{ex,rad}$ and $W^{ex,ang}$. Since the contour is not aligned $w_{ij}^{ex,ang}$ approaches 0 for all ij (see eq. 4.10 and figure 4.13). Given that $w_{ij}^{ex,ang}$ approaches 0, w_{ij}^{ex} also approaches 0. Thus, when the model is presented with stimuli with these characteristics, W^{inh} dominates the connection strengths. It is also important to note that (on average), contour-to-background and background-to-contour connection strengths approach 0 as the difference in SF preference between contour and background neural units becomes larger (because of the effects of W^f - see eq. 4.9 and figure 4.16 panel **A**). With a 5 oct. difference in SF preference between these two groups, (on average) connection strengths are close to 0 (see mean connection strength). It is also worth mentioning that in the right figure of panel **B**, contour-to-contour connections show the highest inhibitory interactions due to the RFs of these neural units being in close proximity to one another. In all other cases (on average) the level of inhibition was lower because there are many distant connections, and W^{inh} decays with distance.

Panel **C** (of figure 4.17) shows the combination of the effects described for panels **A** and **B** (i.e.: for an oriented contour, and for a contour defined by an SF difference between contour and background, respectively). Excitatory interactions are seen for contour-to-contour connections due to the alignment of the contour. Inhibitory interactions dominate connections of neural units which are in close proximity and are not align. This is the case because excitatory interactions are weak between unaligned units, and inhibitory interactions are strong (when the RF of the units are close). The (mean) strength of contour-to-background, and background-to-contour, connections is reduced in comparison to that seen in panel **A** due to the effects of W^f. As the difference in SF preference between neural units ij increases, w_{ij}^f decrease for these connections (see eq. 4.9 and figure 4.8, panel **B**). Since values in W^f decrease, the magnitude of the product of W^f with $\left(w_0^{ex}W^{ex}(E) - w_0^{inh}W^{inh}(E)\right)$ (see eq. 4.4) also decreases. In the case being described (panel **C**), the difference in SF preference between contour and background neural units is high. Thus, the magnitude of the product mentioned above is low, leading contour-to-background and background-to-contour connection strengths to approach 0.

Finally, panel **D**, in figure 4.17, shows (the average) W when the model is presented with stimuli with no contours embedded in them. Inhibitory connections dominate between neural units which have RFs in close proximity, and connection strengths approaching 0 are seen for units which have distant RFs. No excitatory connections are seen because contour elements were randomly aligned, thus $w_{ij}^{ex,ang} \to 0$ for all ij. As described for panel **B**, what would be contour-to-contour connections[4] have the strongest inhibitory connections because the RFs for these neural units are in close proximity to one another.

[4]These are referred to as "would be" because there was no contour defined in these stimuli. However, would be contour units represent neural units with RFs at the location where a contour element would be present if a contour was defined.

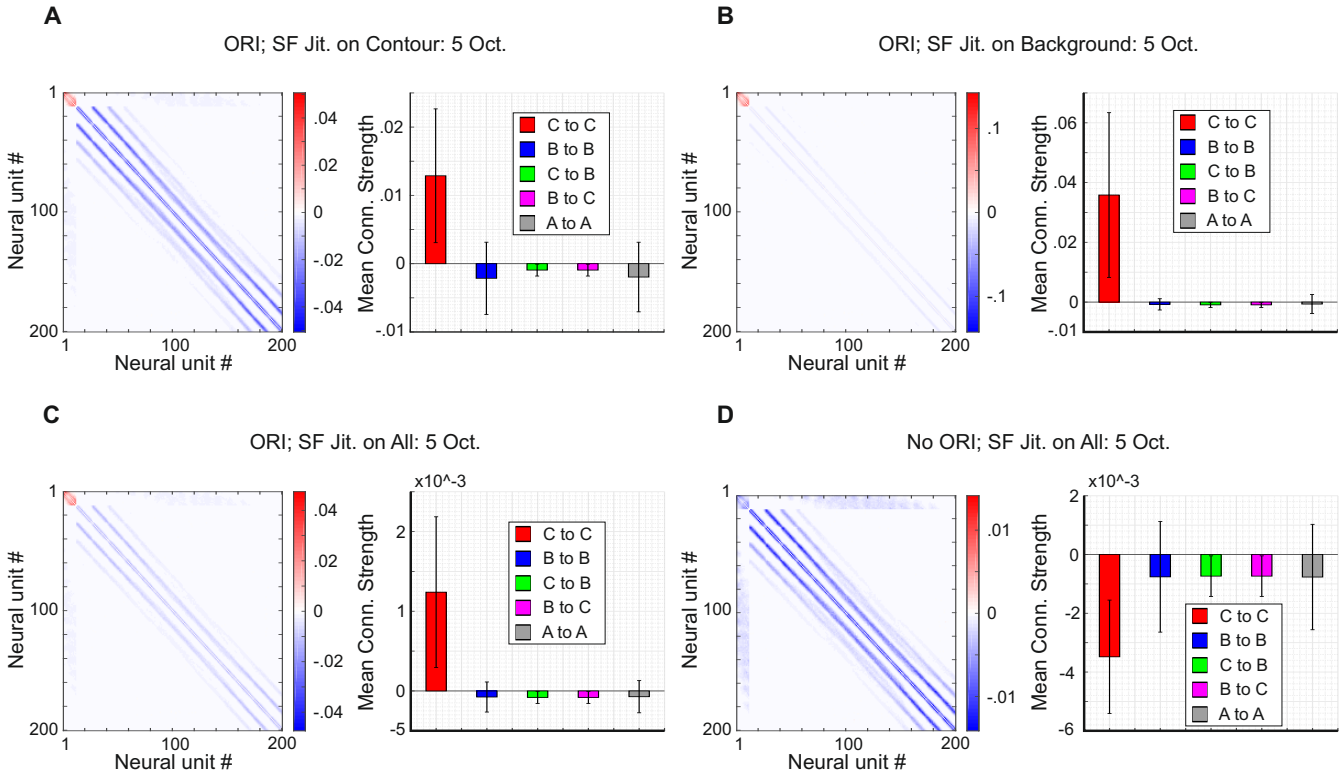

Figure 4.18: Visualization of W and mean connection strengths for different clusters of neural units (as indicated in the plots). **A**: Oriented contour, a homogeneous SF preference for background units, and a SF preferences sampled from a 5 oct. wide normal distribution for contour units. **B**: Oriented contour, a homogeneous SF preference for contour units, and SF preferences sampled from a 5 oct. wide normal distribution for background units. **C**: Oriented contour, and SF preferences sampled from a 5 oct. wide normal distribution for all neural units. **D**: Random orientation for all elements, and SF preferences sampled from a 5 oct. wide normal distribution for all neural units. All other conventions are the same as in figure 4.17.

Figure 4.18 uses the same type of displays as figure 4.17 to show W's properties for Es with different characteristics. In all panels, except for panel **D**, the data corresponds to aligned contour stimuli with varying SF preferences for neural units. Since the effects of an aligned contour in W have already been described, in the following paragraphs, only the effects of SF jitters will be discussed.

Data in panel **A** shows the average W for cases in which the SF preference of background units was homogeneous, and that of contour units was sampled from a (5 oct. wide)[5] uniform distribution. In comparison with a model stimulated with an aligned contour and a homogeneous SF preference for all neural units (figure 4.17, panel **A**), the main differences seen here are the following: **a.** a reduction in (mean) contour-to-contour connection strengths; and **b.** a small increase in (mean) contour-to-background and background-to-contour connection strengths. These effects are due

[5]The width of the normal distribution is specified in brackets because it is not of out-most relevance. What is of relevance is that the distribution chosen was a normal distribution with a none zero standard deviation.

to the influence of W^f. As shown in figure 4.16 panel **C**, W^f couplings for background-to-background connections are all at a maximum. However, couplings between contour neural units, and contour and background neural units are reduced. Thus, in W, contour-to-contour excitatory couplings are reduced by the reductions in magnitude of W^f, and so is the magnitude of the inhibitory couplings between contour-to-background and background-to-contour connections.

In panel **B** data shows the average W for cases in which the SF preference of contour units was homogeneous, and that of background units was sampled from a (5 oct. wide) normal distribution. In this case, the most notable effect of W^f was the increase in connection strength for all connections which involved a background neural unit (due to a decrease in the magnitude of inhibitory connections). This is expected, as W^f was at its maximum for contour-to-contour neural units, but (on average) was reduced for all other connections (see figure 4.16 panel **B**).

In panel **C** data shows the average W for cases in which the SF preference of all neural units was sampled from a (5 oct. wide) uniform distribution. In this case, on average, the components of W modulated by W^f were reduced in magnitude. Figure 4.16 panel **D** shows that most coupling strengths in W^f are not at the possible maximum of 1 when the model is stimulated with this SF condition. Thus, (on average) in comparison with a case in which all neural units have the same SF preference, $W(E)$ is reduced by the modulation of W^f (i.e.: $W(E) < \left(w_0^{ex} W^{ex}(E) - w_0^{inh} W^{inh}(E)\right)$ - see eq. 4.4 for the definition of $W(E)$).

Finally, in panel **D** data shows the average W for cases in which contour elements were not aligned, and the SF preference of all neural units was sampled from a (5 oct. wide) uniform distribution. As explained above, when the SF preference of neural units is sampled in the manner described, W^f has the effect of reducing the magnitude of all couplings (when compared to a case in which neural units have a homogeneous SF preference). Since the model was presented with stimuli which contained no aligned contour, (mean) connection strengths are inhibitory. This is because W^{ex} approached 0, since $w_{ij}^{ex,ang}$ approached 0 for all ij. As discussed previously, when the model is presented with stimuli which contain no aligned contour, the would be contour-to-contour connections cluster has (on average) stronger inhibitory connections than all other clusters. This is the case because the RFs of contour units are in close proximity, and W^{inh} decays with increasing distances between the RFs of neural units.

4.5 Model's Output and Readout Mechanisms

The model outlined in section 4.4.1 defines the temporal evolution of activities for neural units with a receptive field center located at (x_i, y_i), an orientation preference of ϕ_i, and spatial frequency preference of f_i, for $i = 1...N$. The activity of the neural units in the model can be use to generate an activity map, which depicts the response of each of the neural units at any given time throughout the time of stimulation. The activity can be interpreted as a measure of saliency, with higher values representing a higher saliency. This is a commonly used approach in literature (see e.g.: Hansen and Neumann (2008); Li (1998, 1999)). In a neural context, the activity of the neural units can be interpreted as the firing rate of a neural population in a cortical orientation

column.

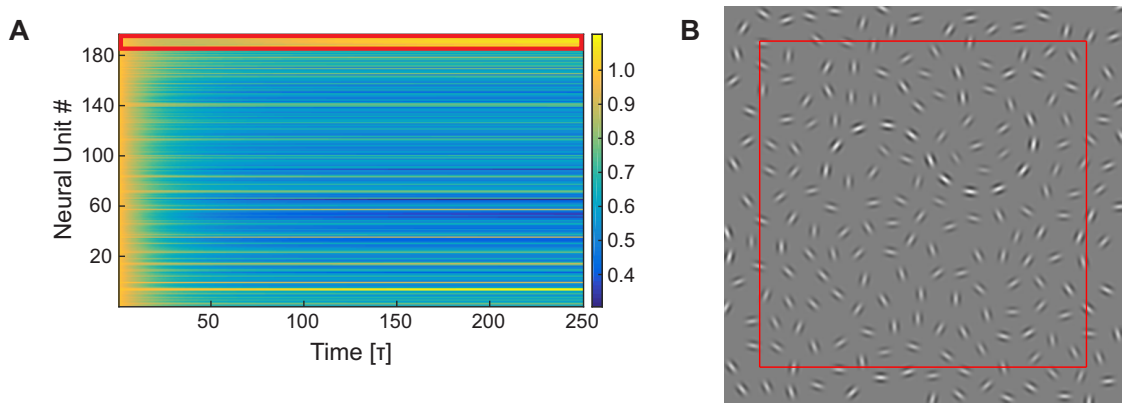

Figure 4.19: **A**: Example of the temporal evolution (in units of τ) of the neural activity levels of the network. Contour units are enclosed in the red rectangle. **B**: Example of a stimulus with contrast proportional to the activity of neural units exited by the edge elements in the display. The activity used to generate the contrast weights was the activity level at the final time step of the stimulation ($\tau = 250$ in panel **A**). The area of the red rectangle represents the area in which line segments, and thus the locations in which the RFs had to be in, in order to be considered for evaluation (see section 4.5.2 for clarification). The data for panels **A** and **B** was obtained using the parameters listed in section 4.4.4.

As can be seen in figure 4.19 after a relatively short period of time (for the case depicted $250 * \tau$, with a $\tau = 25ms$) the network reaches a steady state. Typically the system was stimulated for $50s$ (with $\tau = 25ms$) which provided plenty of time for the network to reach a steady state.

4.5.1 Readout Mechanisms

In order to evaluate the model's performance in 2AFC contour integration task, the output has to be interpreted through a readout mechanism. The main aim when developing a readout mechanism is to evaluate the statistics of a saliency map, in order to select either an area, or in this case a stimulus (i.e.: E), with a particularly noteworthy saliency. To achieve this goal, seven readout mechanisms were envisioned:

$$k_{sum} = \Theta\left(\left(\sum_{i=1}^{N^{Tar}} A_i^t\right) - \left(\sum_{i=1}^{N^{Dis}} A_i^d\right)\right) \tag{4.16}$$

$$k_{mean} = \Theta\left(\left(\frac{1}{N^{Tar}}\sum_{i=1}^{N^{Tar}} A_i^t\right) - \left(\frac{1}{N^{Dis}}\sum_{i=1}^{N^{Dis}} A_i^d\right)\right) \tag{4.17}$$

$$k_{square} = \Theta\left(\left(\frac{1}{N^{Tar}}\sum_{i=1}^{N^{Tar}} \left(A_i^t\right)^2\right) - \left(\frac{1}{N^{Dis}}\sum_{i=1}^{N^{Dis}} \left(A_i^d\right)^2\right)\right) \tag{4.18}$$

$$k_{max} = \Theta\left(\sum_{i=1}^{N^{Tar}} \left(A_i^t\right)^{\infty} - \sum_{i=1}^{N^{Dis}} \left(A_i^d\right)^{\infty}\right) \tag{4.19}$$

$$s^2 = \Theta\left(\frac{1}{N^{Tar}-1}\sum_{i=1}^{N^{Tar}}\left(A_i^t - \bar{A}^t\right)^2 - \frac{1}{N^{Dis}-1}\sum_{i=1}^{N^{Dis}}\left(A_i^d - \bar{A}^d\right)^2\right) \tag{4.20}$$

$$s^3 = \Theta\left(\frac{1}{N^{Tar}-1}\sum_{i=1}^{N^{Tar}}\left(\frac{A_i^d - \bar{A}^d}{\sigma}\right)^3 - \frac{1}{N^{Dis}-1}\sum_{i=1}^{N^{Dis}}\left(\frac{A_i^t - \bar{A}^t}{\sigma}\right)^3\right) \tag{4.21}$$

$$s^4 = \Theta\left(\frac{1}{N^{Tar}-1}\sum_{i=1}^{N^{Tar}}\left(\frac{A_i^d - \bar{A}^d}{\sigma}\right)^4 - \frac{1}{N^{Dis}-1}\sum_{i=1}^{N^{Dis}}\left(\frac{A_i^t - \bar{A}^t}{\sigma}\right)^4\right) \tag{4.22}$$

N^{Tar} and N^{Dis} refer to the number of neural units modelled when the model was presented with a target stimulus, or with a distractor stimulus, respectively. Θ refers to a Heaviside function, which ensures that the estimators return an output of 1 when the first term is larger than the second, and a 0 when the second term is larger than the first. A_i^t refers to the activity of neural unit i in a target trial, when the neural activity has reached a fixed point. Similarly, A_i^d refer to the activity of neural unit i in a distractor trial, when the neural activity has reached a fixed point.

Neural units which do not have a similar mean number of neighbours (similar to most others in the display), because they are close to the display's borders, may have atypical activities. The atypical activities can be due to these units not being influenced by surrounding elements to the same extent as neural units found close to the center of the display. As such, only the activities of neural units with an RF center located in a frame of width of 1.5° of visual angle from the borders of the stimulus were considered. Thus, $A^{t,d} \in$ Inside Containing Rectangle (see figure 4.19, panel **B**). After applying this selection, only elements which were roughly about 15.5° of visual angle from the center of the display were used in the evaluation of the readout mechanisms. It was ensured that the target contour would always be located in a containing box with boundaries of 11° of visual angle form the center of the screen. Thus, the contour elements were always part of the subset used in the evaluation of the readout mechanisms.

4.5.2 Computing Performance

The readout mechanisms can be applied to the neural activity given by N^{Trials}, which consist of a target stimulus ensemble E^t, and a distractor stimulus ensemble E^d - see figure 4.6. The output of the readout mechanisms will either be a 1 (in the case that the target trial was more "noteworthy" than the distractor trial - noteworthy in terms of the neural activities observed once a fixed point was reached), or a 0, in the opposite

case. By applying these readout mechanisms to a large number of trials, and thus obtaining a result for the 2AFC between the target and distractor trial, a contour detection performance score can be computed. The model's performance in contour detection is thus defined as:

$$\boldsymbol{P}_M = \frac{1}{N^{Trials}} \sum_{i=1}^{N^{Trials}} \rho_i \tag{4.23}$$

where ρ_i is the binary output of any of the readout mechanisms for trial i.

All reported model performances in this thesis are based on $N^{Trials} = 1000$, unless otherwise stated. It was ensured that the positions of edge elements in E^t and E^d were not exactly the same (but that they had similar statistics), since this would have led to artificially high performances. If the positions of edge elements in E^t and E^d are the same, then the differences in activities given by the model would only be driven by the angular (see equation 4.10) and the spatial frequency (see equation 4.9) parts of the association field, and not by the radial (see equation 4.6) and inhibitory parts (see equation 4.8). As a consequence of this measure, the sum estimator (see eq. 4.16) had to be transformed into a mean estimator (see eq. 4.17), since the number of neural units in different E^t and E^d was not always the same.

In the 2AFC tasks used in the psychophysical experiments of interest (see section 4.2), target and distractor stimuli were also realized with different positions for the edge elements. Thus, the technique used to compute contour detection performance scores in the modeling experiments, is compatible with what was done in the psychophysical experiments.

A confidence interval (for all $\boldsymbol{P}_M$) was obtained through binomial statistics. For a detailed description of how these were calculated please refer to Appendix C.1. Note that in later sections of this chapter there will be frequent references to the confidence intervals obtained for different $\boldsymbol{P}_M$.

4.6 Evaluation of The Model

The principal aim when devising this model was to create a simplistic, and biophysically plausible, model of feature integration which: **a.** is sensitive to alignment and spatial frequency (dis)similarities; and **b.** is able to reproduce psychophysical performance results. Achieving this would allow for the study of the interactions between these two features, and to create testable hypotheses on the expected perceptual behaviour of observers when presented with novel stimuli. To that aim, first the quantitative reproduction of key psychophysical results described in section 4.2 will be reported. Then, the performance of the model on well established psychophysical tasks will be reported. Finally, the performance of the model when presented with stimuli that have yet to be use in psychophysical experiments will be reported.

In order to simplify the discussion of experimental tasks in the following sections, the reader should take note of the following abbreviations:

- **ORI**: Orientation alignment only condition. Contours were defined only on the

basis of orientation alignment of the edge elements to a global contour path. No other distinguishing qualities were present in the stimulus.

- **SF**: Spatial frequency shift only condition. Contours were defined only on the basis of a fixed spatial frequency difference between the contour elements' spatial frequency, and that of the background elements. No other distinguishing qualities were present in the stimulus. The spatial frequency difference could take two forms. $\mathbf{SF}_{\uparrow}$, in which the spatial frequency of the elements manipulated was higher than that of the other elements, or $\mathbf{SF}_{\downarrow}$, in which the spatial frequency of the elements manipulated was lower than that of the other elements.

- **ORI and SF**: Combination of orientation alignment and spatial frequency shift condition. Contour were defined on the basis of orientation alignment of the edge elements to a global contour path, and a fixed spatial frequency difference between the contour elements' spatial frequency and that of the background elements. No other distinguishing qualities were present in the stimulus. Similarly to the **SF** condition, the shift of the spatial frequency could also take an upwards or downwards form in relation to the non-manipulated elements. These conditions are abbreviated as **ORI and** $\mathbf{SF}_{\uparrow}$ and **ORI and** $\mathbf{SF}_{\downarrow}$ accordingly.

- **SF JIT on BKG**: Spatial frequency jitter on background elements. The stimulus was composed of an orientation aligned contour, with a fixed spatial frequency for the contour elements, in a field of gabor patches with spatial frequencies which could take any value within a given range (for background elements). No other distinguishing qualities were present in the stimulus.

- **SF JIT on CONT**: Spatial frequency on contour elements. The stimulus was composed of an orientation aligned contour, with spatial frequencies for the contour elements which could take any value within a given range. Background elements had a fixed spatial frequency. No other distinguishing qualities were present in the stimulus.

- **SF JIT on ALL**: Spatial frequency jitter on all elements. Combination of the **SF JIT on BKG** and **SF JIT on CONT** conditions. The stimulus was composed of an orientation aligned contour, in a field of gabor patches with spatial frequencies which could take any value within a given range. Both, contour elements and background elements, were defined with spatial frequencies drawn from the same distribution. No other distinguishing qualities were present in the stimulus.

In all the cases in which the spatial frequency of elements was sampled from a distribution, a normal distribution with a none zero standard deviation was employed.

For the model, when the spatial frequency of elements in a display were manipulated, this implied that the preferred spatial frequency of neural units stimulated by said elements were set to the spatial frequency of the edge elements.

It should also be noted that there only was a difference in performance results between the *Fixed Scaling Model* and the *Variable Scaling Model* when the preferred

spatial frequencies of feature detectors were not homogeneous. As such, only when a stimulus with a manipulation on spatial frequency was used will there be two sets of results reported. One for the *Fixed Scaling Model*, and one for the *Variable Scaling Model*. These will be clearly marked. When no such mention of results pertaining to one of these two specific variants of the model is made, the results apply equally to both model variants.

Lastly, whenever reference is made to the activity of edge elements, or to the stimulation of edge elements, what is meant is the activation (or stimulation), of neural units which are excited by said edge elements.

4.6.1 Primary Validation: Reproduction of Psychophysical Experiments

In the following subsections, performance results of the model in a 2AFC contour detection task (calculated as described in 4.5.2), will be reported for a number of distinct stimulation conditions. For practical reasons, two parameter sets, one for the unidirectional model and one for the bidirectional model, will be used to report all results, unless otherwise stated (see section C.2 for a detailed explanation on how these parameters were selected). k_{mean} (see equation 4.17) was the estimator selected to report all results. The decision to use this estimator was made with consideration to biological plausibility, and by manually comparing the results yielded by this estimator to those yielded by the other estimators. In this comparison, a (manual) qualitative decision was made on which estimator provided the best compromise for the basic results which the model was required to reproduced. These basic results are those outlined in section 4.2.

The selected values for the parameters of the model are as follows:

1. Unidirectional

$$\lambda_f = 1.45 \quad \lambda_{ex} = 0.5 \quad \lambda_{in} = 0.6 \quad w_0^{ex} = 30 \quad w_0^{inh} = 1.8 \quad \sigma_\alpha = 0.2 \quad \sigma_\beta = 0.8$$

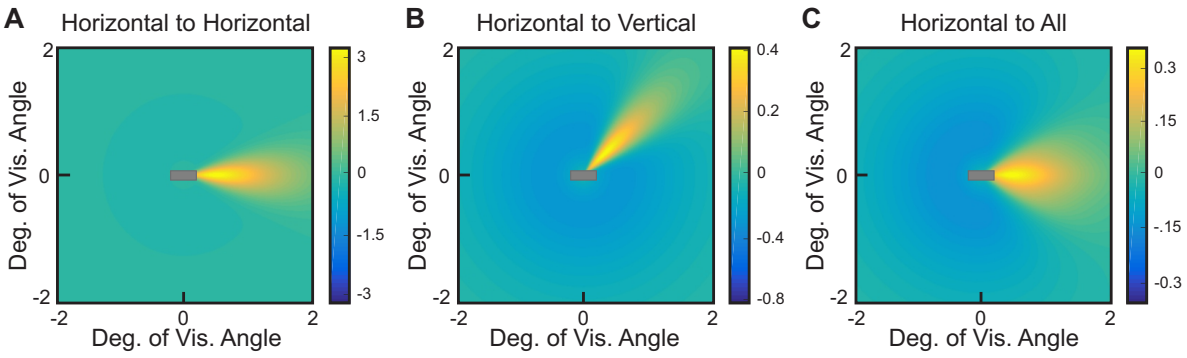

Figure 4.20: Association field (W - eq. 4.4) for the unidirectional model with no spatial frequency differences between elements. **A**: Horizontal to horizontal coupling strengths. **B**: Horizontal to vertical coupling strengths. **C**: Horizontal to all orientations coupling strengths. All plots use an edge element (shown by the grey line in the center of the display), which stimulates a neural population with an RF centered at that visuotopic location, as a presynaptic origin.

2. Bidirectional

$$\lambda_f = 1.15 \quad \lambda_{ex} = 0.5 \quad \lambda_{in} = 0.6 \quad w_0^{ex} = 10 \quad w_0^{inh} = 1 \quad \sigma_\alpha = 0.1 \quad \sigma_\beta = 0.8$$

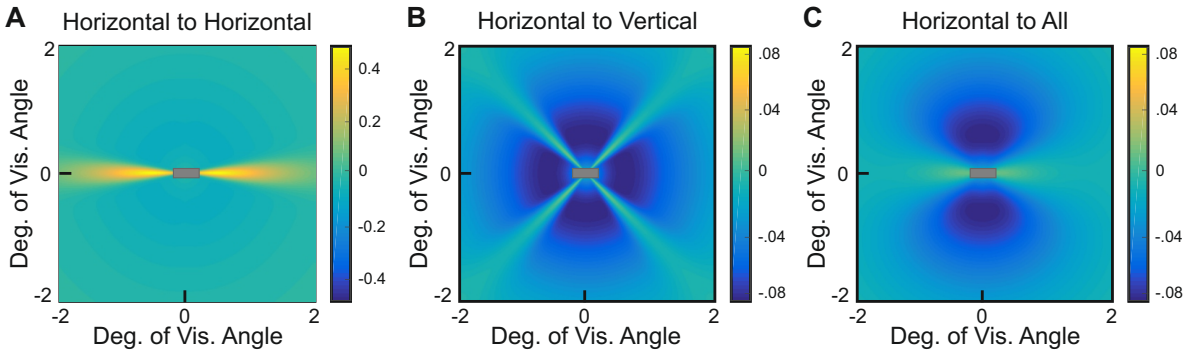

Figure 4.21: Association field (W - eq. 4.4) for the bidirectional model with no spatial frequency differences between elements. **A**: Horizontal to horizontal coupling strengths. **B**: Horizontal to vertical coupling strengths. **C**: Horizontal to all orientations coupling strengths. All plots use an edge element (shown by the grey line in the center of the display), which stimulates a neural population with an RF centered at that visuotopic location, as a presynaptic origin.

Given the large number of experimental conditions to which the model was subjected, it was not possible to achieve an exact quantitative reproduction of all psychophysical results with a unique parameter set. However, most results were qualitatively reproduced by the parameter sets selected. In the few instances when a reproduction of results was not achieved, the possible reasons regarding the lack of success will be discussed.

4.6.1.1 Orientation Only Condition

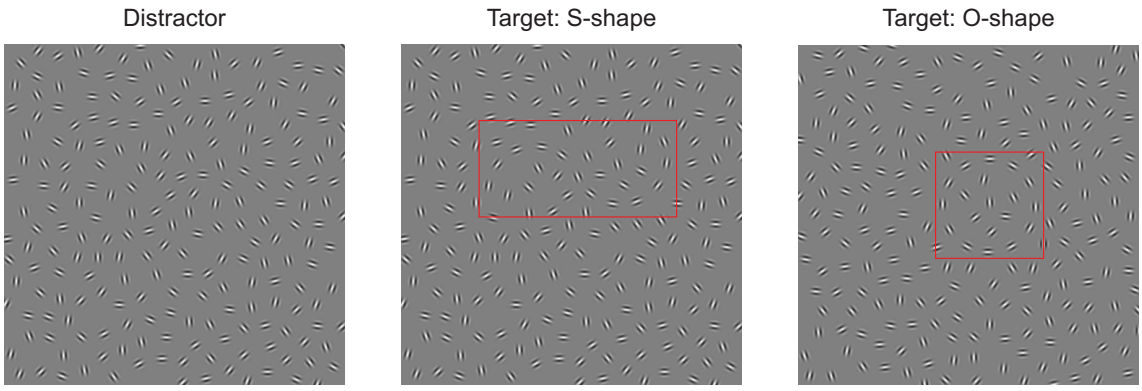

Figure 4.22: Examples of the stimuli used. The red rectangles in the two target stimuli were added to help the reader find the contour.

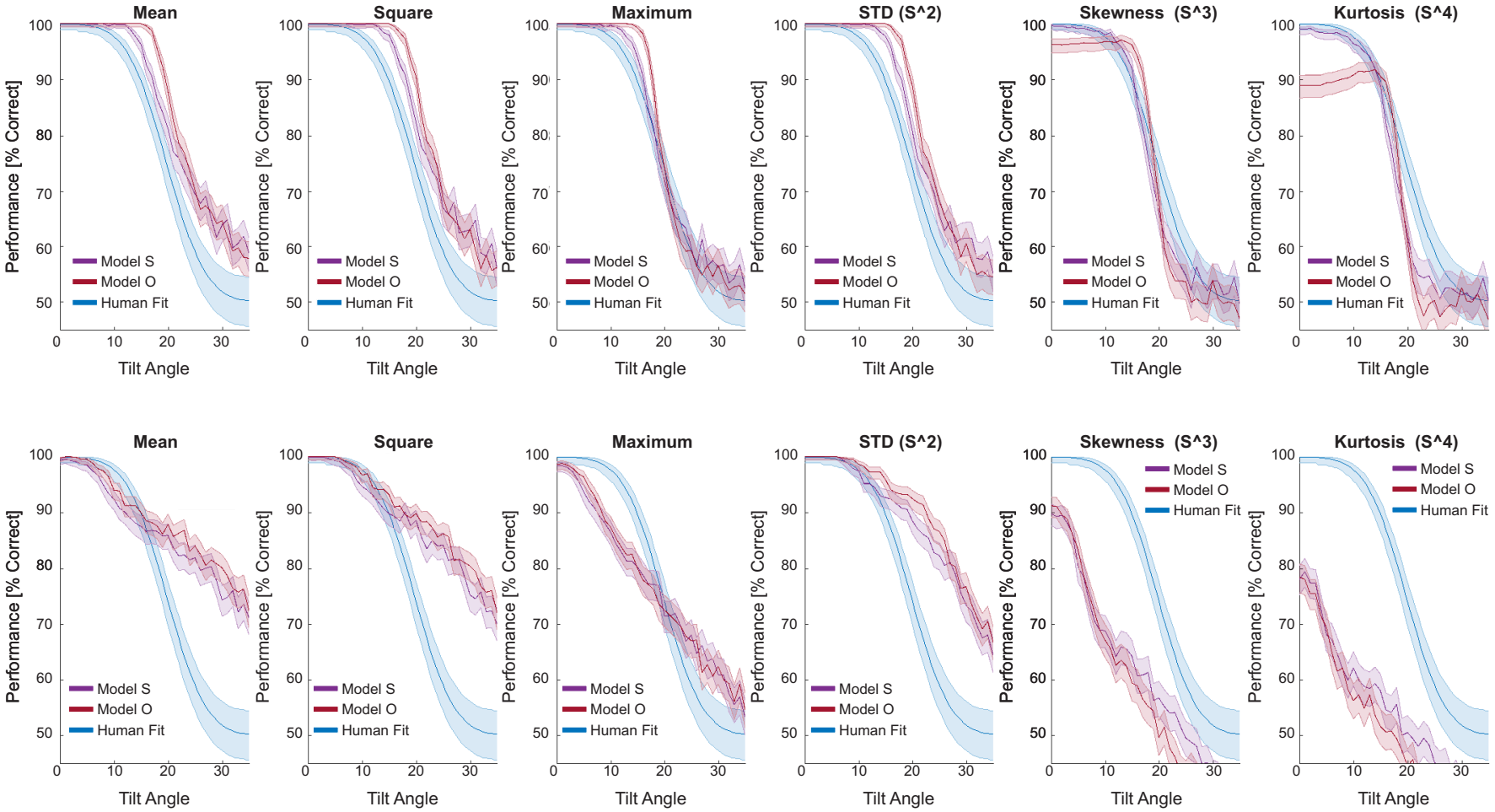

Figure 4.23: Performance for the six estimators as indicated by the panel's title. Top six panels: results for the unidirectional model. Bottom six panels: results for the bidirectional model. Blue curve represents the fitted data of the human performance with S-shape stimuli (Persike and Meinhardt, 2015b); Burgundy curve represents the model's performance with O-shape stimuli; Purple curve represents the model's performance with the S-shape stimuli. The shaded areas represent the expected measurement errors for the corresponding curves according to binomial statistics.

One of the key qualitative results of contour integration literature reported by Persike and Meinhardt (2015a,b); Persike et al. (2009) (and many other - e.g.: Field et al. (1993); Hess and Field (1999); Kovacs (1996)) is that contour detection performance decreases with increased deviations of contour elements' alignment to the path of a contour (from now on this deviation will be referred to as tilt angle jitter). As such, this was the first result which the model was required to reproduce. Both S-shaped contours and O-shaped contours were used as stimuli. Figure 4.22 shows examples of both types of stimuli.

As can be seen in figure 4.23 performance varied drastically depending on the readout mechanisms applied. With the exception of s^3 and s^4 (see section 4.5.1) all readout mechanisms applied to the unidirectional model yielded performance results which were better than human performance for the selected parameter set. Only k_{mean}, k_{square} and s^2 yielded equivalent (or better) results to human performance with the bidirectional model, for the parameters selected. However, as can be seen in figure 4.23 all estimators for both models were able to reproduce the results qualitatively

4.6.1.2 Spatial Frequency Only Condition

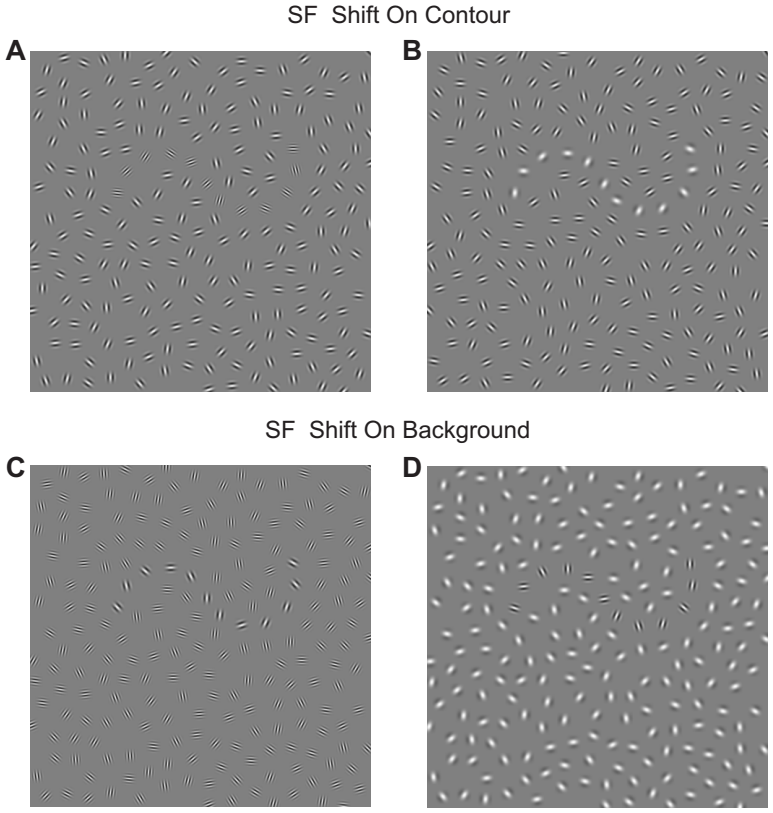

Figure 4.24: Spatial frequency shift example stimuli. Stimuli used for the psychophysical experiments and to stimulate the model. The spatial frequency of either contour, or background elements (as indicated in the image), was shifted upwards (**A** and **C**) or downwards (**B** and **D**) with respect to f_0.

In a set of two experiments in which the spatial frequency (SF) of contour elements was manipulated in order to differ from the SF of background elements, Persike and Meinhardt (2015a) showed that contour detection based solely on SF differences between these two classes of elements is possible. As can be seen in figure 4.24, contours are visible when they are defined by spatial frequency differences between contour and background elements. These type of stimuli were used in the psychophysical experiments and also to stimulate the model.

Persike and Meinhardt (2015a) showed that: **a.** contour detection performance is dependent on the magnitude of the difference between the spatial frequencies of contour and background elements; and **b.** that octave shifts of the same magnitude are more effective (i.e.: result in a bigger performance increase) if the shift is done in the negative direction ($\mathbf{SF}_{\downarrow}$ condition), rather than in the positive direction ($\mathbf{SF}_{\uparrow}$ condition).

Fixed Scaling

As seen in figure 4.25, the *Fixed Scaling Model* is able to deliver the whole spectrum of performance results for the 2AFC task (i.e.: performances between 0.5 and 1 proportion correct).

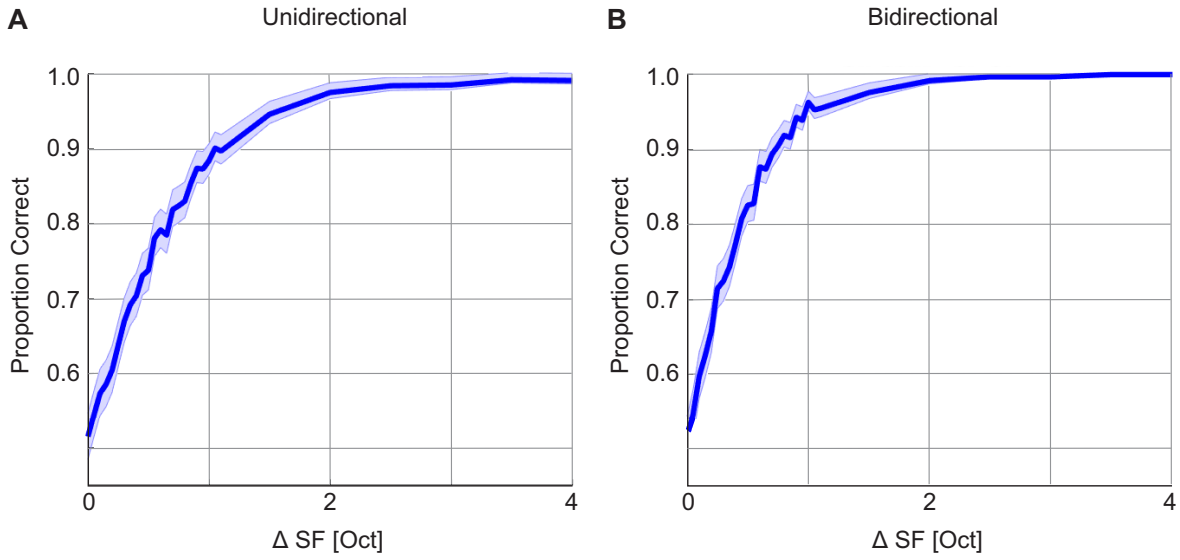

Figure 4.25: Performance results for spatial frequency shifts (*Fixed Scaling Model*). **A**: Unidirectional model; **B**: Bidirectional model. Shaded areas represent the expected error on the measurements based on binomial statistics.

In the *Fixed Scaling Model*, no distinction is made between an upward, or a downward, SF shift (nor between whether it is contour units, or background units, that have an SF preference other than f_0). These distinctions are not relevant in the *Fixed Scaling Model* because the only component of this model which is directly affected by SF is W^f (see equation 4.9). In equation 4.9 absolute values are used. Thus, the result of $|f_i - f_j|$ is not affected by whether f_i is larger than f_j, or vice-versa.

Variable Scaling

Figures 4.26 and 4.27 show the results of the unidirectional and bidirectional models (respectively), for an $\mathbf{SF}_{\uparrow}$ and $\mathbf{SF}_{\downarrow}$ manipulation on either contour or background (as indicated in the figures).

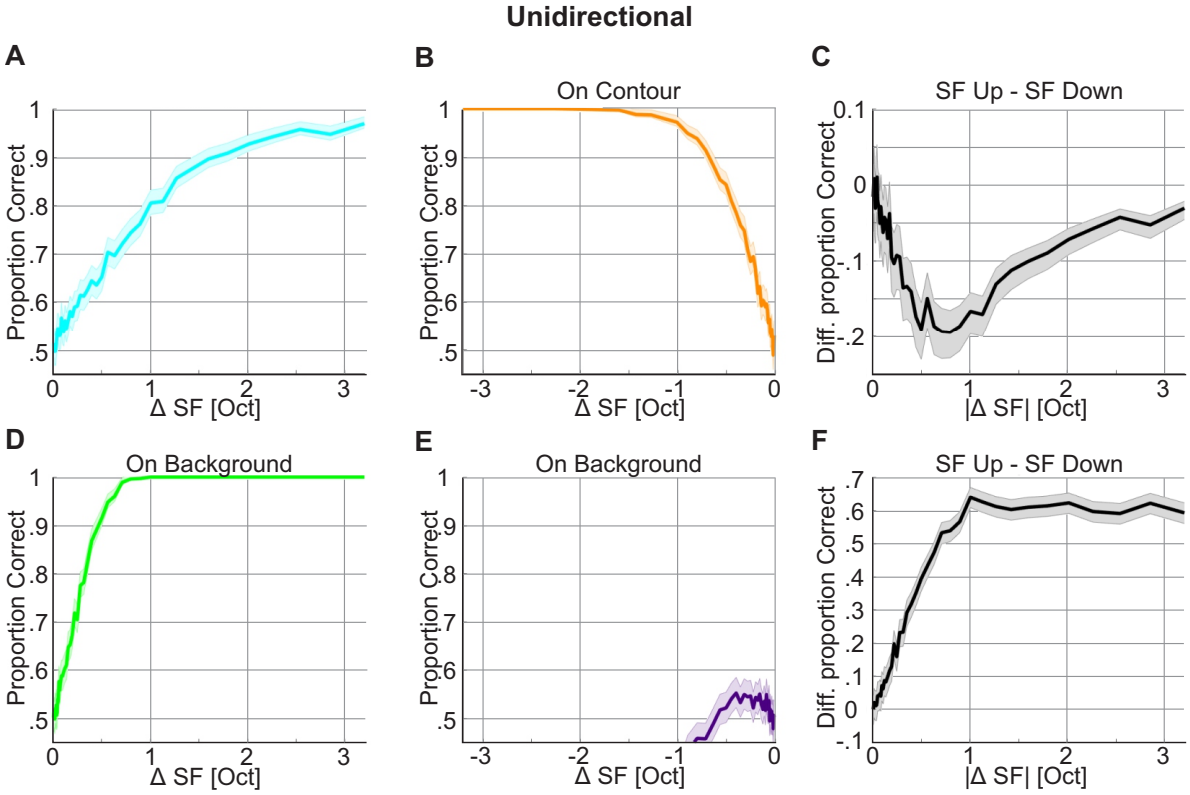

Figure 4.26: Performance results for the unidirectional model stimulated with four spatial frequency shift conditions. **A**: $\mathbf{SF}_{\uparrow}$ on contour; **B**: $\mathbf{SF}_{\downarrow}$ on contour; **C**: Difference in performance gains between the $\mathbf{SF}_{\uparrow}$ and $\mathbf{SF}_{\downarrow}$ on contour conditions. Positive values indicate better performance for the $\mathbf{SF}_{\uparrow}$ condition. Panels **D, E** and **F** represent the same stimulation conditions as panels **A, B** and **C**, with the only difference that the source of variation in spatial frequency were background elements rather than contour elements. Shaded areas represent the expected error on the measurements based on binomial statistics.

The unidirectional and bidirectional models produced psychometric curves ranging from 0.5 to 1 proportion correct for all conditions, except for the $\mathbf{SF}_{\downarrow}$ on background condition (panel **E**). In this condition there was an initial small increase in performance, followed by a drop in performance that went below chance level (at approximately -1 to -1.5 Δ SF, depending on the model's directionality). Performance went below chance level since neural activities in distractor trials were higher than in target trials. When comparing selected coupling clusters (see discussion below) in W (eq. 4.4)[6], it

[6]For a W obtained when presenting the model with $\mathbf{SF}_{\downarrow}$ on background stimuli, or with $\mathbf{SF}_{\uparrow}$ on background stimuli

becomes evident why performance decreased in the $\mathbf{SF}_{\downarrow}$ on background condition, and why it increased in the $\mathbf{SF}_{\uparrow}$ on background condition. In the $\mathbf{SF}_{\uparrow}$ and $\mathbf{SF}_{\downarrow}$ on contour conditions, there was a difference in the effectiveness of the SF manipulations when the SF preference was higher ($\mathbf{SF}_{\uparrow}$ condition), or lower ($\mathbf{SF}_{\downarrow}$ condition) than f_0. As seen in panel **C** of figures 4.26 and 4.27, the $\mathbf{SF}_{\downarrow}$ manipulation was more effective than the $\mathbf{SF}_{\uparrow}$ manipulation (before performance reached 100% correct). A discussion of these two remakes ensues.

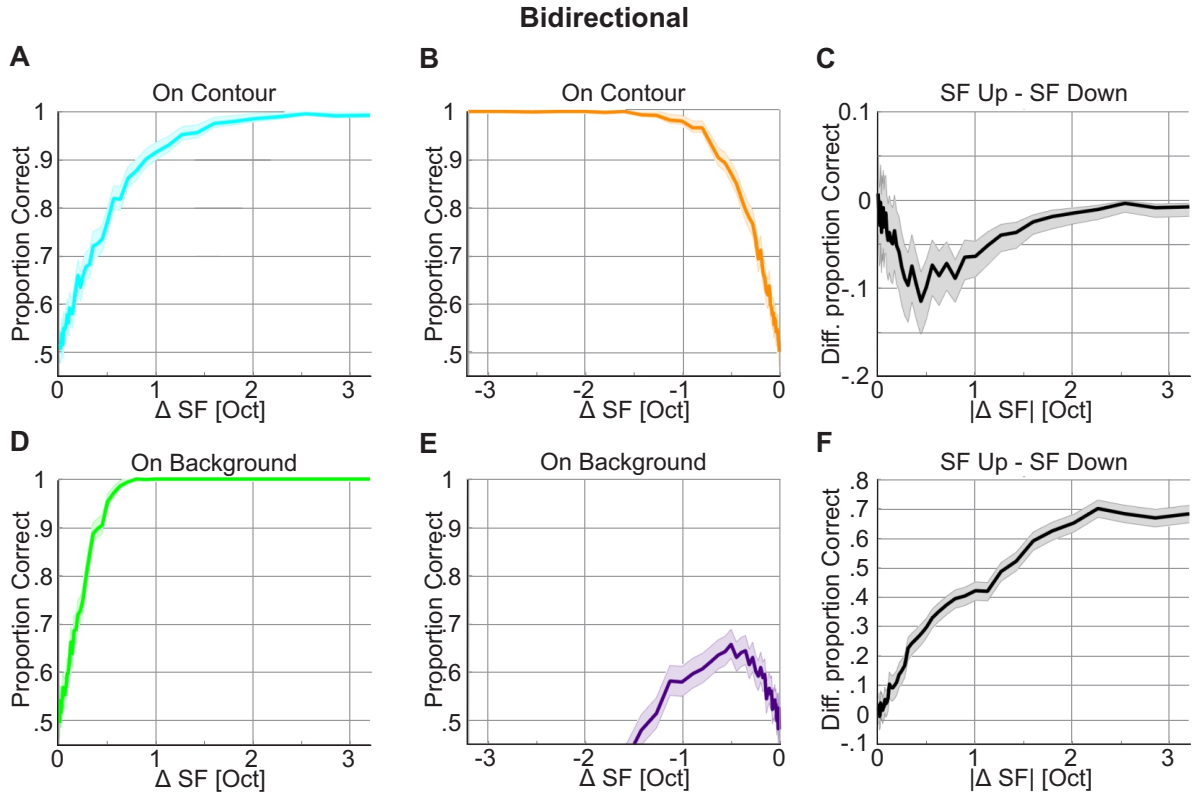

Figure 4.27: Performance results for the bidirectional model stimulated with four spatial frequency shift conditions. Same conventions as in figure 4.26.

In order to facilitate this discussion please note that the connection clusters (in W) to be compared are the following: contour-to-contour, background-to-background, contour-to-background, and background-to-contour. These connection clusters will be referred to as C to C, B to B, C to B, and B to C, respectively.

1. ***Effects of an SF on Background Manipulation***:

   Figure 4.28 shows different plots outlining the changes in the average connection strengths of all the clusters mentioned above. In panel **A** of this figure, data pertaining to the $\mathbf{SF}_{\downarrow}$ on background condition is displayed. In panel **B**, data pertaining to the $\mathbf{SF}_{\uparrow}$ on background condition is shown.

   Performance decrease in $\mathbf{SF}_{\downarrow}$ on background: when stimulating the model with **distractor** stimuli (with the $\mathbf{SF}_{\downarrow}$ on background manipulation), connection strengths for neural activities increased (blue curve, panel **A**). Distractor stimuli

consisted of ensembles without an oriented contour, and with a homogeneous SF preference for all units which was smaller than f_0. Since the SF preference of all units was homogeneous, w_{ij}^{f} equaled 1 for all ij (see eq. 4.9). However, the magnitude of $\lambda_{ex_{ij}}$, increased due to the SF preference of neural units being lower than f_0 (see eq. 4.7). This increase in $\lambda_{ex_{ij}}$ extended the reach of lateral interactions in $w_{ij}^{ex,rad}$, leading to higher $w_{ij}^{ex,rad}$ values for all ij, and ultimately to stronger excitatory couplings (see figure 4.15, panel **C** for an illustration of how excitatory couplings change with changes in λ_{ex}).

When stimulating the model with **target** stimuli (red curve), in the $\mathbf{SF}_{\downarrow}$ on background condition, there was no increase in connection strengths for the C to C cluster (see figure 4.28, panel **A**). Furthermore, the increase in the connection strengths of the B to C, and C to B clusters, were not as high when presenting the model with target stimuli, as they were when presenting the model with distractor stimuli. C to C connection strengths remained constant when presenting the model with **target** stimuli because neither W^f, nor λ_{ij} (the only two components directly affected by SF manipulations), were affected in this condition. B to C and C to B connection strengths increased primarily because the magnitude of w_{ij}^{f} decreased for these connections. Since these connections are mainly inhibitory, a decrease in w_{ij}^{f} lead to stronger connection strengths for the relevant ij (see figure 4.17, panel **C** and its associated discussion for details). B to C connection strengths reached higher levels, and had a higher rate of increase than C to B connections, because the length scale of excitatory interactions (i.e.: λ_{ex}) increased for these connections. This increase lead to longer lateral interaction for excitatory connections, which lead to an increase in $w_{ij}^{ex,rad}$, and in turn to an increase in w_{ij}^{ex} for these connections (see figures 4.14 and 4.15 panel **C** for an illustration on how $W^{ex,rad}$ and W^{ex}, respectively, change with an increase λ_{ex}).

Lastly, panel **A** of figure 4.28, shows that in the B to C and C to B plots, the blue and the red lines intersect approximately -1.5 Δ SF. This explains why performance is seen to go below chance level in the $\mathbf{SF}_{\downarrow}$ on background condition at approximately the same point (see figure 4.27, panel **E**). At this point the connection strengths of B to C and C to B become larger in distractor stimuli than in target stimuli. Thus, when connection matrices with these properties are used as recurrent input for the model, neural activity is higher for distractor trials than it is for target trials (after integrating the system until it reaches a steady state).

Performance increase in $\mathbf{SF}_{\uparrow}$ on background: unlike in the $\mathbf{SF}_{\downarrow}$ on background condition, performance increased in the $\mathbf{SF}_{\uparrow}$ on background condition. This is because connection strengths increased in B to C and C to B connections when the model was presented with target stimuli, and they decreased when the model was presented with distractor stimuli (as seen in figure 4.28, panel **B**). The increase in B to C, and C to B, connection strengths was mainly due to the effects of W^f, which reduced the magnitude of inhibitory connections (see figure 4.17, panel **C**, and its associated discussion for details).

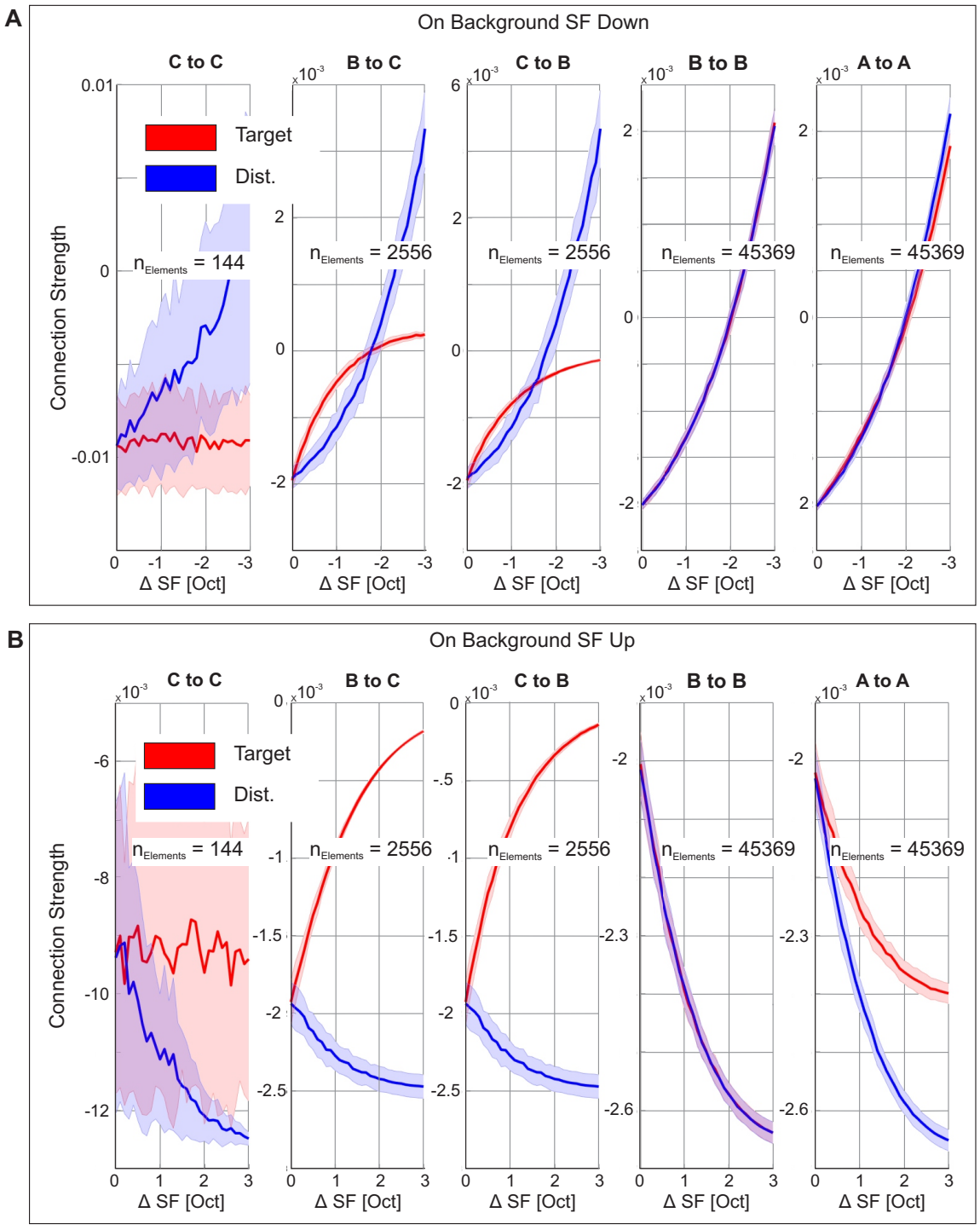

Figure 4.28: Mean connection strengths between neural units, based on varying SF shifts from f_0 in background elements. 1000 weight matrices were used to obtain the statistics. **C to C**: contour to contour connections; **B to C**: background to contour connections; **C to B**: contour to background connections; **B to B**: to background connections; **A to A**: all to all connections. **A**: $\mathbf{SF}_{\downarrow}$ on background units; **B**: $\mathbf{SF}_{\uparrow}$ on background units elements. Shaded areas represent standard deviations. $n_{Elements}$ refers to the number of connections. Thus, the number over which mean connection strengths were obtained for each of the different groups, in each of the association fields. The bidirectional model was used to obtained the weight matrices (W).

When presenting the model with distractor trials, W^f did not have an effect because the SF preference of all neural units was homogeneous. The decrease in all connection strengths was due to the effects of a decrease in the magnitude of $\lambda_{ex_{ij}}$ for all ij. A decrease in magnitude of $\lambda_{ex_{ij}}$ leads a decrease in the reach of W^{ex}'s lateral interactions (see figure 4.15, panel **C**). Thus, with shorter lateral interactions in W^{ex}, the amount of excitation a neural unit receives decreases. Since excitatory connections become weaker as Δ SF increases, connection strengths become weaker for all types of neural unit connections.

2. ***Effects of an SF on Contour Manipulation***:

When the SF preference of contour units was deviated from f_0, the $\mathbf{SF}_\downarrow$ condition was more effective than the $\mathbf{SF}_\uparrow$ condition (see panel **C** in figures 4.26 and 4.27). The $\mathbf{SF}_\downarrow$ manipulation yielded a faster rate of improvement in contour detection performance, up to (approximately) 2Δ SF, when both the $\mathbf{SF}_\downarrow$ and $\mathbf{SF}_\uparrow$ reach 100% correct performance. The greater effectiveness of the $\mathbf{SF}_\downarrow$ on contour condition was due to the effect of the variable reach of lateral interactions in W^{ex} (realized as defined in eq. 4.7). As seen in panel **C** of figure 4.15, when the SF preference of neural units is lower than f_0, the reach of lateral interactions in W^{ex} increases. This increase in lateral interactions leads to stronger excitatory couplings. On the other hand, when the SF preference of neural units is higher than f_0, the reach of lateral interactions in W^{ex} decreases (see panel **C** of figure 4.15). This decrease in the reach of lateral interactions leads to weaker excitatory couplings in the $\mathbf{SF}_\uparrow$ than in the $\mathbf{SF}_\downarrow$ (on contour) condition. Thus, these changes in w_{ij}^{ex}, lead to greater contour detection performance in the $\mathbf{SF}_\downarrow$ than in the $\mathbf{SF}_\uparrow$ on contour condition (with the same magnitude of Δ SF, before 100% correct detection performance is reached).

Finally, note that unlike in the $\mathbf{SF}_\downarrow$ on background condition, both, the $\mathbf{SF}_\uparrow$ and $\mathbf{SF}_\downarrow$ (on contour) conditions yielded psychometric curves which ranged between 50 and 100% correct. The psychometric curves for these conditions were never below chance level because of the nature of target and distractor stimuli. Target stimuli were composed of contours defined only on the basis of an SF difference between contour and background elements (i.e.: contour elements were not aligned), and distractor stimuli consisted of randomly oriented elements with a homogeneous SF preference of f_0. Distractor trials in the SF on background, and in the SF on contour conditions, differed in one important aspect. In the SF on contour conditions, the SF preference of neural units was always f_0, thus $\lambda_{ex_{ij}}$ was λ_{ex_0} for all ij. In the SF on background condition, the SF preference for all neural units was that which differed from f_0 in the corresponding target trials. Thus, in the distractor trials of the SF on background conditions $\lambda_{ex_{ij}}$ was the same for all ij, but not equal to λ_{ex_0}.

As seen in figure 4.29, when the model is presented with these type of stimuli, B to C, and C to B connections are weaker for distractor stimuli (SF Homo.), than they are for contour stimuli (SF Up and SF Down). These connections are weaker because of the effects of W^f (see figure 4.17, panel **B** and its associated discussion for details). C to C connection strengths are weaker for the $\mathbf{SF}_\downarrow$,

and stronger for the $\mathbf{SF}_{\uparrow}$ (on contour) condition, because of the effects of $\lambda_{ex_{ij}}$ in $W^{ex,rad}$. In the $\mathbf{SF}_{\downarrow}$ (on contour) condition the reach of $w_{ij}^{ex,rad}$ is extended for connections originating from a contour unit, thus extending the lateral reach of excitatory connections, which leads to stronger connections (when the source of connection is a contour neural unit). The opposite effect is true for the $\mathbf{SF}_{\uparrow}$ (on contour) condition. In this condition the reach of $w_{ij}^{ex,rad}$ is shortened for connections originating from a contour unit, thus shortening the lateral reach of excitatory connections, which leads to weaker connections when the source of connection is a contour neural unit.

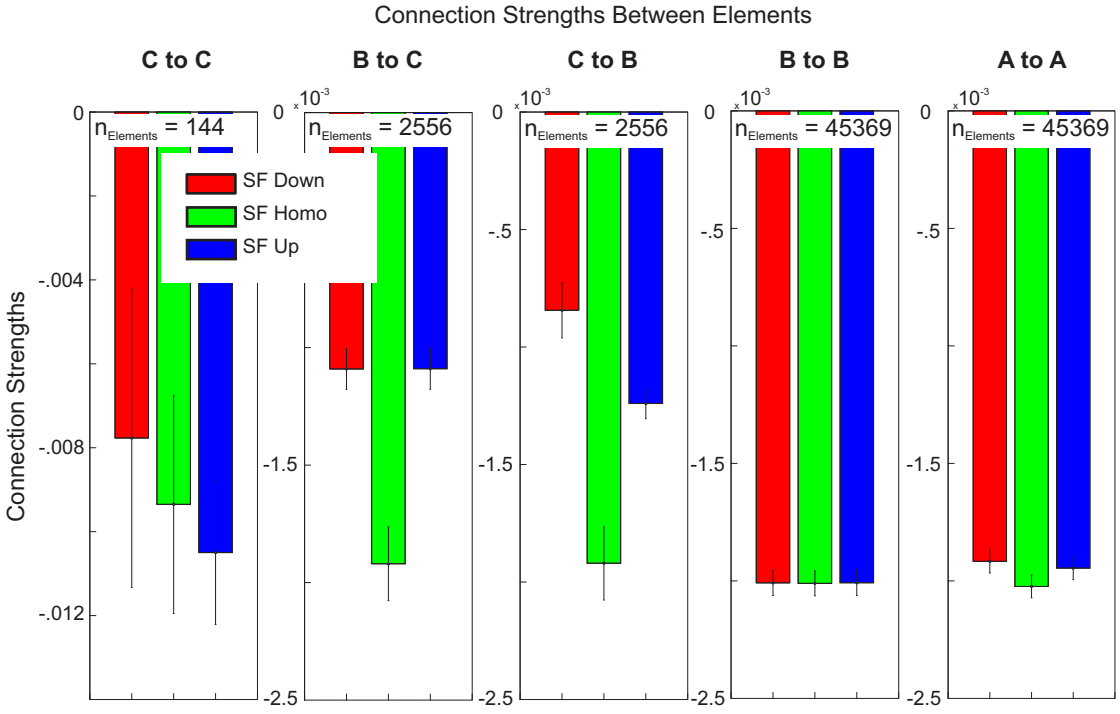

Figure 4.29: Mean connection strengths of neural units when the SF preference of contour units are manipulated. 1000 weight matrices were used to obtain the statistics. Red: 0.65 octaves lower than f_0 SF preference; Green: f_0 SF preference; Blue: 0.65 octaves higher than f_0 SF preference.

4.6.1.3 Feature Contrast Levels

In their experiment Persike and Meinhardt (2015a) determined the feature contrast levels (i.e.: SF magnitude differences between contour and background elements) required to achieve four different performance levels (dubbed visibility levels). The four performance levels were 62%, 68%, 74% and 80% correct performance. In figures 4.30 and 4.31 the visibility levels results for the *Fixed Scaling Model* and the *Variable Scaling Model* are shown, respectively, along side the results from Persike and Meinhardt (2015a).

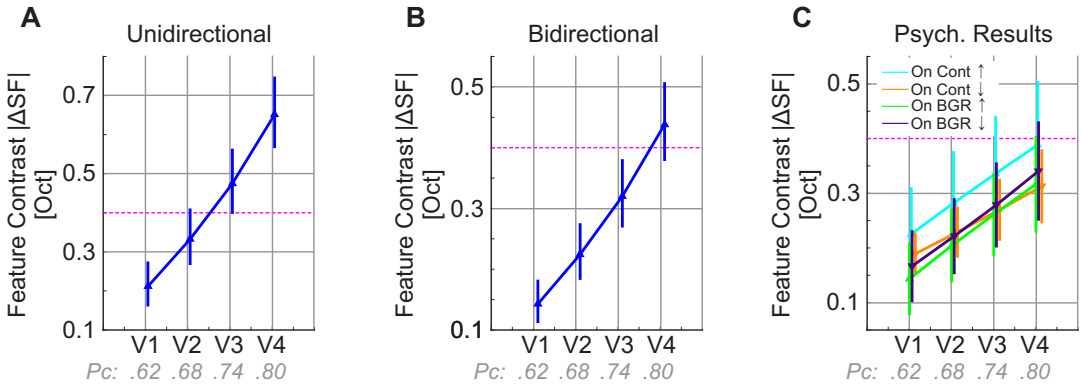

Figure 4.30: Feature contrast levels at the four visibility levels - *Fixed Scaling Model*. **A:** Unidirectional Model; **B:** Bidirectional Model; **C:** Human data (Persike and Meinhardt, 2015a). Vertical bars indicate expected error on the measurements based on binomial statistics for panel **A** & **B** and standard deviation for panel **C**. The magenta line marks the 0.4 feature contrast $|\Delta SF|$ for comparison purposes. In panels **A** and **B** a single line is presented as all **SF** conditions are identical in the *Fixed Scaling Model*.

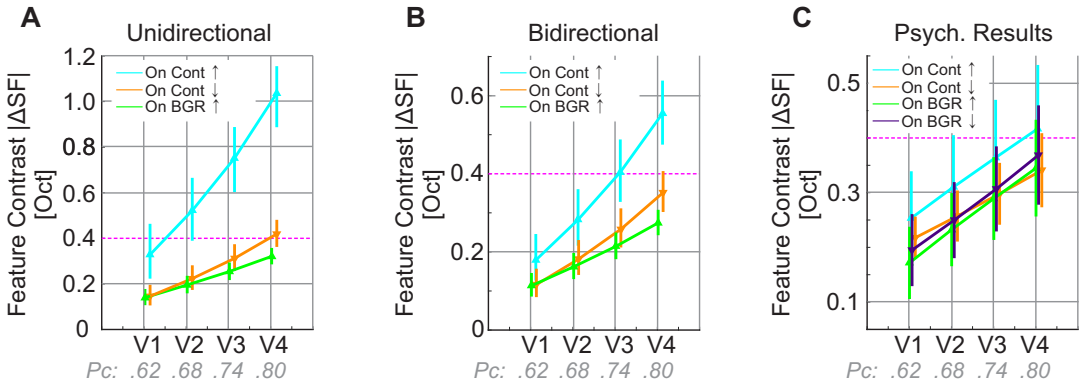

Figure 4.31: Feature contrast levels at the four visibility levels - *Variable Scaling Model*. All other conventions as in figure 4.30.

As can be seen for the *Fixed* and *Variable* scaling model results, the bidirectional model was able to reach the four distinct visibility levels with smaller feature contrast levels. However, for all model results a greater magnitude in feature contrast levels was needed to reach the visibility levels than in the psychophysics experiments.

Variable Scaling

There is a clear slope difference between the conditions in the model results. This implies that not only are smaller feature contrast levels needed to achieve the same performance in the different conditions, but also that smaller differences between the contrast levels are needed to attain a higher visibility level in the $\mathbf{SF}_{\downarrow}$ on contour and the $\mathbf{SF}_{\uparrow}$ on background conditions, than in the $\mathbf{SF}_{\uparrow}$ on contour condition.

The pattern of results is not as clear in the psychophysical data (panel **C** of figure 4.31), since the data has a large standard deviation over feature contrast. On average, in the $\mathbf{SF}_{\uparrow}$ on contour condition the largest feature contrast was required to attain each of the visibility levels, as was the case in the model. The $\mathbf{SF}_{\uparrow}$ on background and $\mathbf{SF}_{\downarrow}$ on contour conditions yielded performances too similar to one another, thus, it is hard to make a strong statement on which required a greater feature contrast to achieve each of the visibility levels.

As discussed in section 4.6.1.2, and as seen in panels **A** and **B** (of figure 4.31), to effect the same changes in performance, different magnitudes of feature contrast were needed in the $\mathbf{SF}_{\uparrow}$ and $\mathbf{SF}_{\downarrow}$ conditions. The source of this difference are the effects of $\lambda_{ex_{ij}}$ in $W^{ex,rad}$. In $\mathbf{SF}_{\uparrow}$ condition the lateral reach of W^{ex} was shortened (for the neural units with a lower SF preference than f_0), leading to weaker excitatory connections (in which the source unit was one of the affected units by the change in SF preference). In $\mathbf{SF}_{\downarrow}$ condition the lateral reach of w_{ij}^{ex} was extended for the pertinent ij, leading to stronger excitatory connections (see figure 4.15 panel **C**).

4.6.1.4 Orientation and Spatial Frequency Conditions

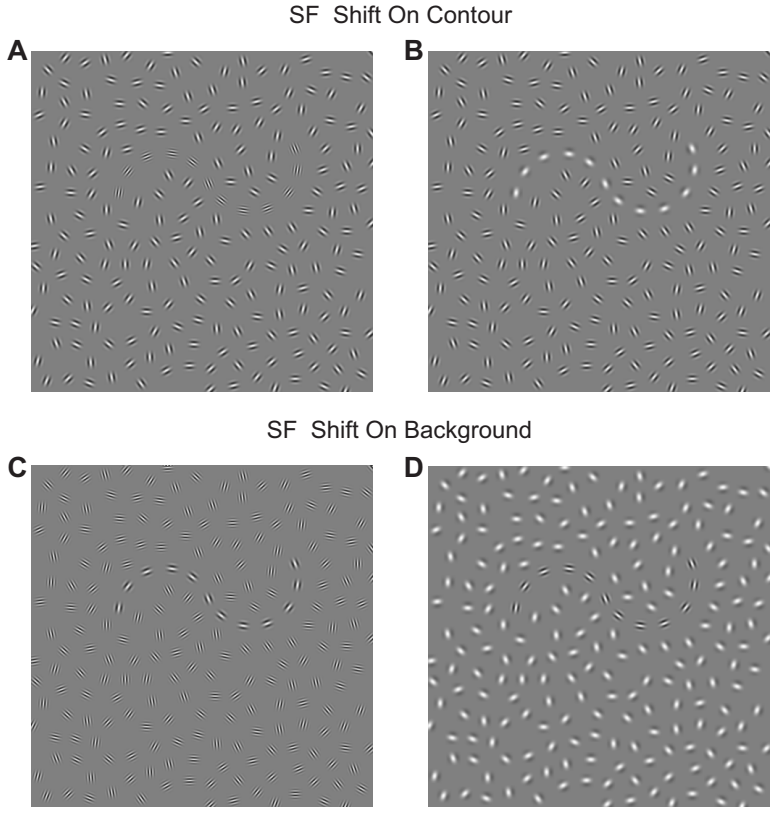

Figure 4.32: Examples stimuli. The spatial frequency of either contour, or background elements (as indicated in the image), was shifted upwards (**A** and **C**) or downwards (**B** and **D**) with respect to f_0. Note that in these stimuli contour elements are aligned to the contour path, as opposed those in figure 4.24.

Once the four visibility thresholds for the **ORI** and different **SF** conditions were found by Persike and Meinhardt (2015a), they proceeded to combine cues gieven by these two features. Four conditions were created: **ORI and SF**$_{\uparrow}$, **ORI and SF**$_{\downarrow}$ on contour, and **ORI and SF**$_{\uparrow}$, **ORI and SF**$_{\downarrow}$ on background. Examples of these type of stimuli can be seen in figure 4.32.

In order to do simulations replicating these conditions, the feature levels which led to performances of 62, 68, 74 and 80% correct in each of the conditions with a single feature were found (see figures 4.30 and 4.31 for visibility levels in the **SF** conditions). For the *Fixed Scaling Model* the single features were **ORI** only and **SF** shift, and for the *Variable Scaling Model* the single features were **ORI** only, **SF**$_{\uparrow}$ only, **SF**$_{\downarrow}$ only (applied to either contour or background elements as appropriate).

Fixed Scaling

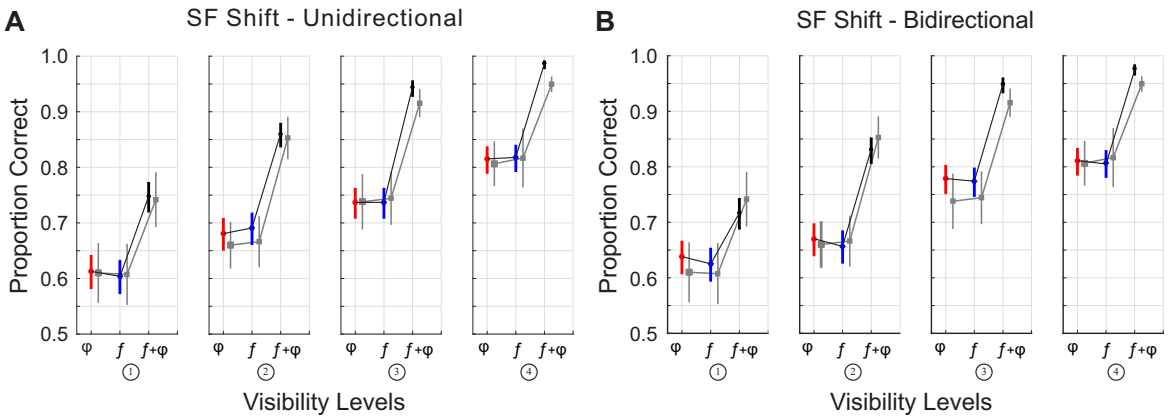

Figure 4.33: Performance for **ORI** (red - labeled as ϕ), **SF** (blue - labeled as f), and **ORI and SF** (black - labeled as $\phi + f$). Data in grey represents psychophysical data, all other data is related to the model (*Fixed Scaling Model*). Grey data for the **SF** condition represents the average of the four variations (i.e.: on contour/on background, up/down). The four graphs in each of the panels show the four visibility levels, as labeled. **A:** Unidirectional model; **B:** Bidirectional model. For model data, vertical bars represent the expected error on the measurements, based on binomial statistics. For psychophysical data, vertical bars represent standard deviation on the measured data.

The model was stimulated with the **ORI** condition and with the **SF** condition at each of the four visibility levels alone, and also in a combined condition (i.e.: **ORI and SF**). The results of these simulations are presented in figure 4.33. As in the psychophysical experiment (results shown in light grey) there was an increase in performance when the two cues were combined. Both, the unidirectional and bidirectional models show this pattern of results. Unlike in the psychophysical experiments correct detection performance reached approximately 100% when the orientation cue and the spatial frequency cue were combined at visibility level 4. The increase in performance

with the combination of the two cues was achieved because the sources of information regarding the existence of a contour originate from two different mechanisms (which can be combined). The "detection" of the contour based on the orientation cue is done by the angular ($W^{ex,ang}$) and radial ($W^{ex,rad}$) parts of the association field, and the "detection" of the contour based on the SF cue is done by W^{f} (see figure 4.17 - and its associated discussion - for a detailed explanation of how W^{f} interacts with W^{ex}, and the other components of the association field).

Variable Scaling

As with the *Fixed Scaling Model*, the *Variable Scaling Model* was presented with the **ORI** condition, and the appropriate spatial frequency conditions (in an individual and combined manner). The resulting combinations were **ORI and SF$_{\uparrow}$**, **ORI and SF$_{\downarrow}$** on contour, and **ORI and SF$_{\uparrow}$** on background condition. The **ORI and SF$_{\downarrow}$** on background was omitted for reasons previously discussed (see section 4.6.1.2). The results of these simulations, alongside the corresponding psychophysical results (in grey), are presented in figures 4.34 and 4.35 for the unidirectional model and the bidirectional model, respectively.

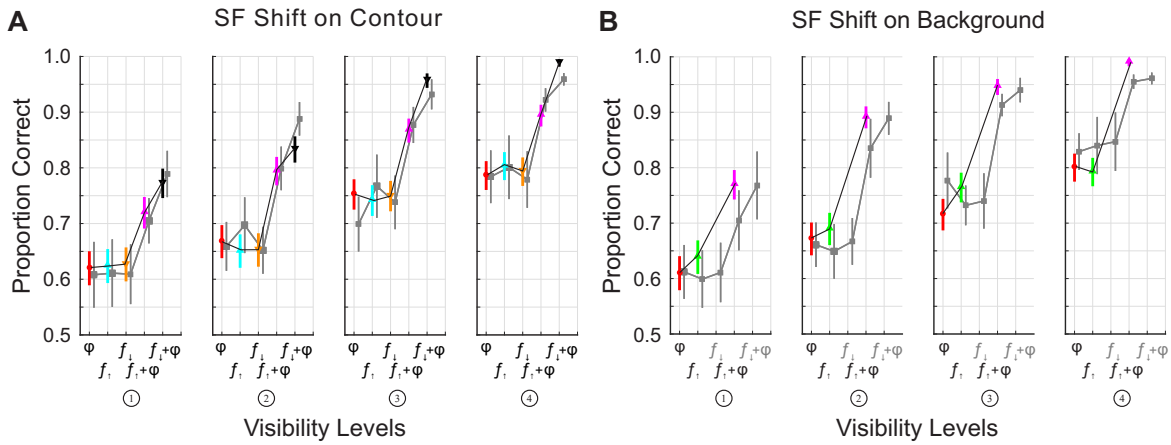

Figure 4.34: Performance for **ORI** , applicable **SF** conditions, and **Ori and** (applicable) **SF** . The colors correspond to those employed in figure 4.31. Data corresponds to the unidirectional model (*Variable Scaling Model*). All other conventions as in figure 4.33. **A SF** on contour **B**: **SF** on background.

It can again be appreciated that the results from the psychophysical experiments performed by Persike and Meinhardt (2015a) were reproduced by the model. The combination of the two features (at all of the visibility levels) lead to a greater performance than either of the individual features (see figures 4.34 and 4.35). Furthermore, in the model results, as in the psychophysical experiments, a clear increase in performance was seen when stimuli composed with the two cues were employed (increase in comparison to the single cues). In the **SF** on contour condition, the mean for correct

detection performance was higher in the **ORI and SF$_\downarrow$** condition than it was in the **ORI and SF$_\uparrow$** condition (panel **A** in both figures - data shown in magenta and black respectively), also as in the psychophysical experiment.

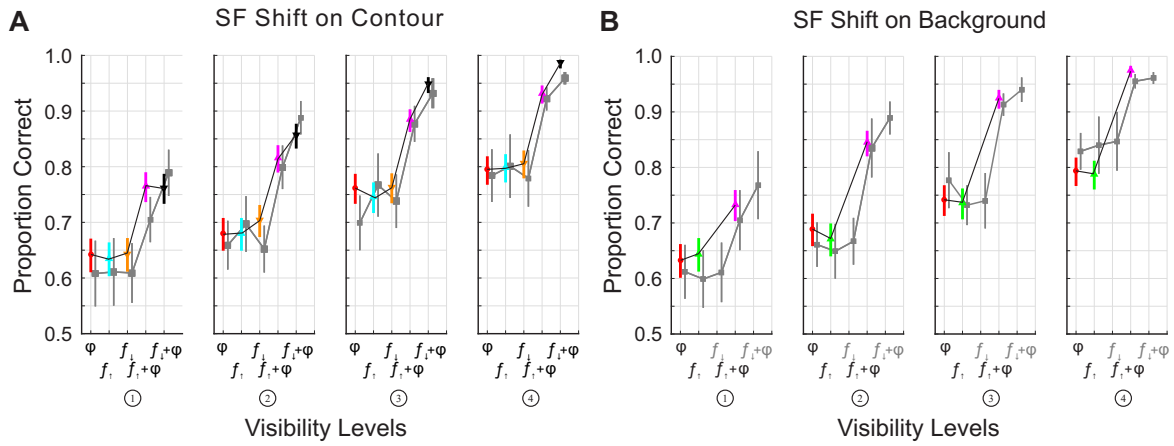

Figure 4.35: Data corresponds to the bidirectional model. Same conventions as in figure 4.34.

Unlike in the *Fixed Scaling Model*, the SF preference of neural units has a direct effect in W^{ex}, by affecting $W^{ex,rad}$. This model can modulate W^{ex} in a unique manner. By having a $\lambda_{ex_{ij}} \neq \lambda_{ex_0}$ (see equation 4.7) for ij pair of neural units in which the SF preference of a source neural unit is affected, differences in the **ORI and SF$_\uparrow$** and **ORI and SF$_\downarrow$** conditions can be achieved.

4.6.1.5 O-Shaped Contours versus S-Shaped Contours

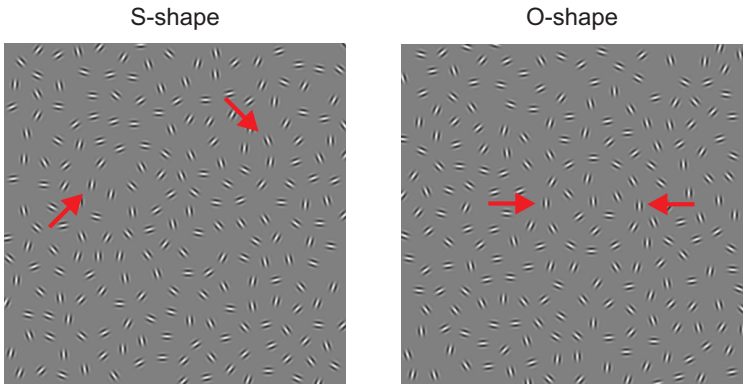

Figure 4.36: Examples of O- and S-shaped target stimuli. Red arrows were added to the figure to mark the location of the contours.

Another key result in psychophysics literature, and also reproduced by Persike and his team, is that contours with an O shape lead to better contour detection performance than S shaped contours (Hess et al., 2001; Kovacs and Julesz, 1993; Mathes and Fahle,

2007; Persike et al., 2009). This can be seen in figure A.2 in the appendix. In the leftmost two panels, on the top and bottom row of this figure, the data shows that subjects could achieve a 75% correct performance with a tilt angle of 19.08° for O-shape contours, and with 14.45° for S-shape contours. Thus, subjects required S-shape contours to have a better alignment to the path angle than O-shape contours to achieve the same performance. See figure 4.36 for examples of O- and S-shaped contours.

The model was able to reproduce this result. As can be seen in figure 4.37 when the model was stimulated with O-shape contours it yielded a higher correct detection performance than when stimulated with S-shape contours (for most tilt angles sampled between 10° and 30°). Thus, as in Persike et al. (2009), S-shape contours had to be better aligned to the path angle than O-shape contours to achieve the same contour detection performance. As seen in panel **A**, between 15 and 20°, there was no overlap in the performance curves yielded by O- and S-shape contours (this also holds true when accounting for measurement errors predicted by binomial statistics). In the range between 10 and 30° of tilt angle, a sampling of 1° was done in order to avoid missing any possible nuances between the performances yielded by the model when stimulated with these two contour shapes.

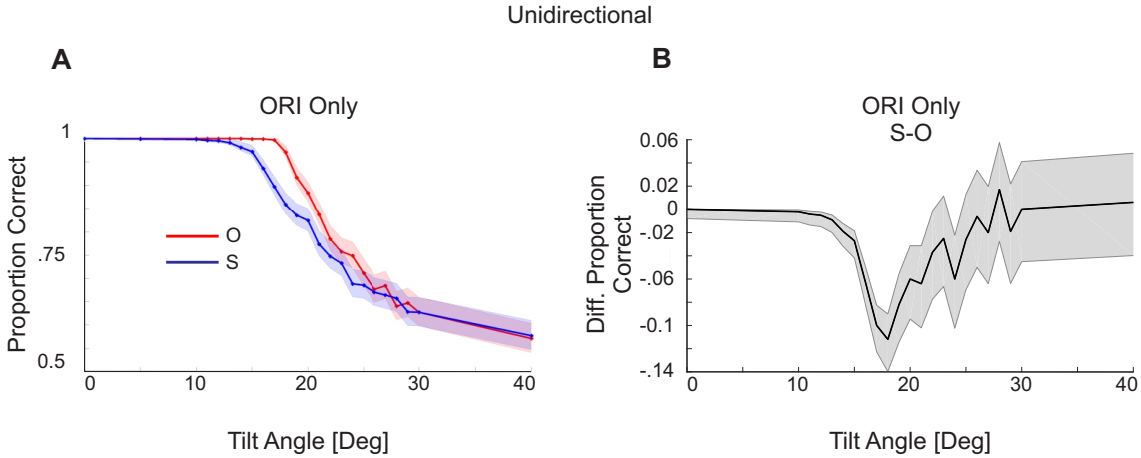

Figure 4.37: **A**: Proportion correct detection performance for contours with O (red) and S (blue) shapes as a function of tilt angle in the contours. Shaded areas represent expected error given by binomial statistics, circular markers represent points sampled. **B**: Difference in proportion correct detection performance for S-O, negative values represent higher proportion correct detection performance for O shape contours. The data corresponds to the unidirectional model.

The 75% performance thresholds for O- and S-shaped contours in the model were higher, in both cases, than the means for the human psychophysical data. The model required 24.5° of tilt angle for O-shaped contours and 23.1° of tilt angle for S-shaped contours to perform at 75% correct (upper and lower bound of the measurement errors predicted by binomial statistics were, respectively, 25.3° and 23.7° for O-shaped contours, and 24.2° and 21.7° for S-shaped contours). In the psychophysics experiment the thresholds found were 19.08° (S.D.: 5.91) for O-shaped contours, and 14.45° (S.D.: 7.45) for S-shaped contours. When taking into account the standard deviation in the

measures obtained from the psychophysical data, the 75% threshold for S-shaped contours is quantitatively reproduced by the model, and the lower bound of the threshold predicted by binomial statistics for O-shaped contours is only 3.4° of tilt angle away from the upper bound of the confidence region in the psychophysical data. Although in the O-shape contours condition the tilt angle required to reproduce the 75% thresholds for the model does not provide an exact quantitative match to the data measured by Persike et al. (2009) an extenuating factor has to be taken into account. The data presented in this section was obtained under a noiseless condition, which is biophysically implausible. As shown in section 4.6.3.2 performance decreases if noise is introduced into the system. Thus with a model which includes noise, the alignment threshold to reach a 75% correct performance will be lower. Indeed, when the simulations were repeated with a noise level of $\sigma^{Noise} = 0.7$, the tilt angle threshold to reach a 75% correct detection performance was 20.1° for O-shaped contours and 18.2° for S-shaped contours, thus providing a good quantitative match when accounting for confidence regions.

As seen in figure 4.38, the bidirectional model also shows a decrease in contour detection performance with increasing tilt angle jitters. The psychometric curve does not have the typical "S" shape because of the σ_α and σ_β chosen for $w^{ex,ang}$ (see figure 4.13 for a comparison on how $w^{ex,ang}$ changes with different choices of σ_α and σ_β, and figure 4.21 for a visualization of the association field used with the bidirectional model).

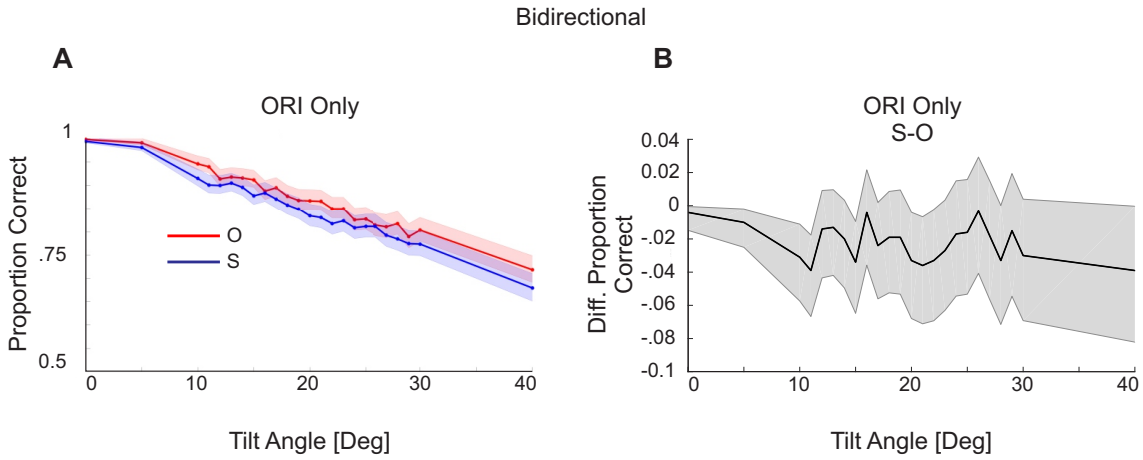

Figure 4.38: Same conventions as in figure 4.37. The data corresponds to the bidirectional model.

4.6.1.6 Jitter on All Elements

Persike and Meinhardt (2015b) showed that by jittering the spatial frequency of all elements in an ensemble, performance decreases. Specifically, when the spatial frequencies of individual gabor elements are sampled from a 2 or 3 octave wide range, contour detection performance decreases (when compared to a condition with homogeneous spatial frequencies for all elements in a stimulus).

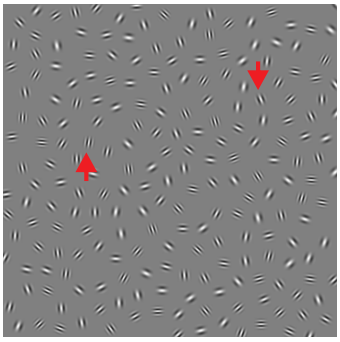

Figure 4.39: Target stimulus with **SF jitter on all elements**

Fixed Scaling

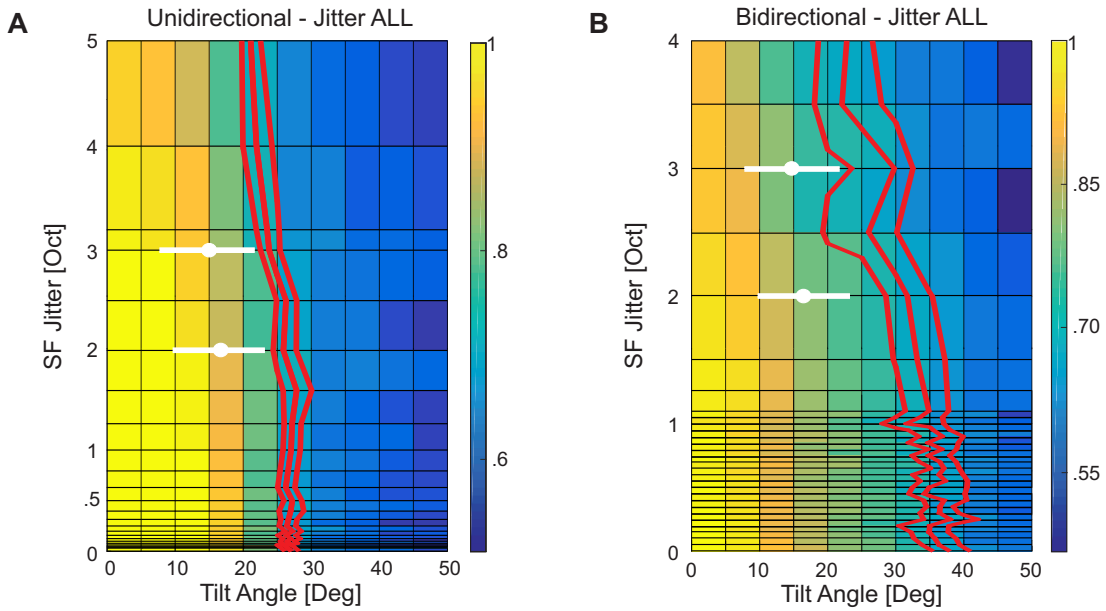

Figure 4.40: Proportion correct matrix for detection performance of contours in a display with SF jitter on all elements, and tilt angle jitters causing deviations from perfect alignment (*Fixed Scaling Model*). **A**: Unidirectional model; **B**: Bidirectional model. Red lines are contour lines at 0.729, 0.75 and 0.776 proportion correct performance (i.e.: 0.75 proportion correct and upper and lower bounds of the expected error given by binomial statistics). White circles represent mean tilt angle threshold for humans to obtain a 75% correct performance at 2 and 3 octave jitters in spatial frequency (Persike and Meinhardt, 2015b), horizontal white lines intersecting the red circles represent standard deviation.

The model was qualitatively able to reproduce this result. As seen in figure 4.40, in both the unidirectional and the bidirectional model results, there is a slight tendency for a lower degree of tilt angle being required to achieve a 75% performance, as spatial frequency jitter increases. This was due to the effects of W^f. A comparison of figure 4.17, panel **A** and figure 4.18 panel **C**, reveals that in the **SF JIT on ALL** condition,

connection strengths are (on average) of a lower magnitude than when all neural units have the same SF preference[7].. As the level of SF jitter increases, the magnitudes of connection strengths decreases. This implies that connection strengths in target and distractor stimuli approach 0 as jitter levels increase in the **SF JIT on ALL** condition. Thus, with higher levels of SF jitter it is harder (at least for the model) to distinguish between a target and a distractor trial.

The effect described occurs because (on average), as the jitter level increases w_{ij}^{f} becomes smaller for all ij. Hence, the effect of increasing levels of jitter, are to decrease connection strengths (see figure 4.18, panel **D** and its discussion for details on how this occurs).

Variable Scaling

The model was able to reproduce Persike and Meinhardt (2015b)'s results. As seen in figure 4.41 with increasing levels of SF jitter, performance decreases. Looking at the 0° tilt angle column, in both panels of figure 4.41, this becomes evident. At low levels of jitter, detection performance is at ceiling (at 100% correct). However, as the level of SF jitter increases, detection performance can be seen to deteriorate. Results also show that with increasing levels of jitter, contour elements need to be better aligned to be detected at the same performance levels. This tendency is shown by the red lines, which represent 75% correct detection performance (and the lower and upper bounds predicted by binomial statistics).

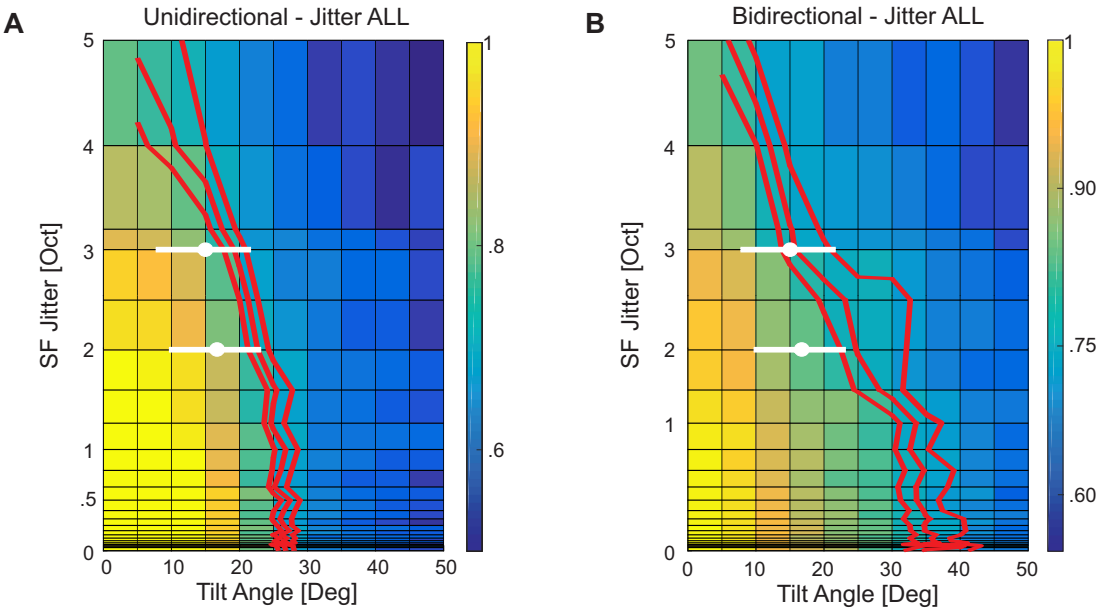

Figure 4.41: All conventions as in figure 4.40 (*Variable Scaling Model*).

[7]This comparison is suggested as the stimuli used to obtain the figures only differ in one aspect. The stimuli used to obtain figure 4.17, panel **A** consisted of an orientation aligned contour and a homogeneous SF for all elements. Whereas, the stimuli used to obtain figure 4.18, panel **D** consisted of an orientation aligned contour and unique spatial frequencies for each element sampled from a uniform distribution.

When taking into account confidence intervals, it can be seen that the tilt angles required to achieve a 75% detection performance (with 2 and 3 octave jitters in SF) are very similar between the psychophysical results and the model results. In the psychophysical experiments at the 2 and 3 octaves jitter levels, a tilt angle of 16.4° and 14.7° (respectively), was required to reach a 75% correct performance level. Both the unidirectional and bidirectional models yielded a better match for the tilt angle threshold at the 3 octave jitters level, than at the 2 octave jitters level. However, confidence regions for the measured human performance and the higher bound of the expected error in the model measurements overlap for the 2 octave jitter level, thus, quantitatively reproducing this result.

As was the case for the *Fixed Scaling Model* results, in the *Variable Scaling Model* results, performance is seen to decrease as jitter levels increase. In the *Variable Scaling Model* results, performance is seen to decrease to a greater extent than in the *Fixed Scaling Model* results (with the same level of SF jitter). Unlike in the *Fixed Scaling Model*, in the *Variable Scaling Model* there are two reasons for the decrease in performance. One is the effect of W^f, which is identical to that described for the *Fixed Scaling Model*. The other is the effect of a $\lambda_{ex_{ij}}$ which is not λ_{ex_0} for all ij. In this model, $\lambda_{ex_{ij}}$ can take any value within a given range, for any ij. That is, there is no structure in which a given cluster[8] of neural units has the same reach of lateral interactions in W^{ex}. This lack of structure in W^{ex}'s lateral interactions reach leads to weaker excitatory connections in W than when all neural units have the same lateral reach.

4.6.1.7 Jitter on Background Elements

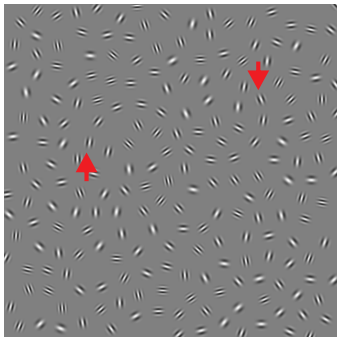

Figure 4.42: Target stimulus with **SF jitter on background elements**

Another condition tested by Persike and Meinhardt (2015b) was that of spatial frequency jitter on background elements. In this condition, the SFs of contour elements were set to the expected value of the distribution from which the SFs of background elements were sampled. In two levels of this condition the spatial frequencies of background elements were sampled from a 2 and 3 octave wide range, respectively.

Results showed that in order to reach a level of 75% correct performance, subjects required the tilt angle of the contour to be 20.4° and 22.2° respectively. This implies

[8]A cluster refers to C to C, B to C, C to B, or B to B connections.

that with increasing jitter levels in the SF of background elements, contours become more visible.

Fixed Scaling

As in the psychophysical experiment results, the tilt angle threshold required to reach a 75% correct contour detection performance remained fairly similar with increasing levels of SF jitter. This was the case for both, the unidirectional and the bidirectional model results (see figure 4.43). The main difference between the unidirectional, and bidirectional, model results is that in the bidirectional model, a higher tilt angle threshold was required to reach a 75% correct contour detection performance. This was the case because the bidirectional model is more resistant to tilt angle manipulations than the unidirectional model (see figure 4.23 - mean estimator for bidirectional model).

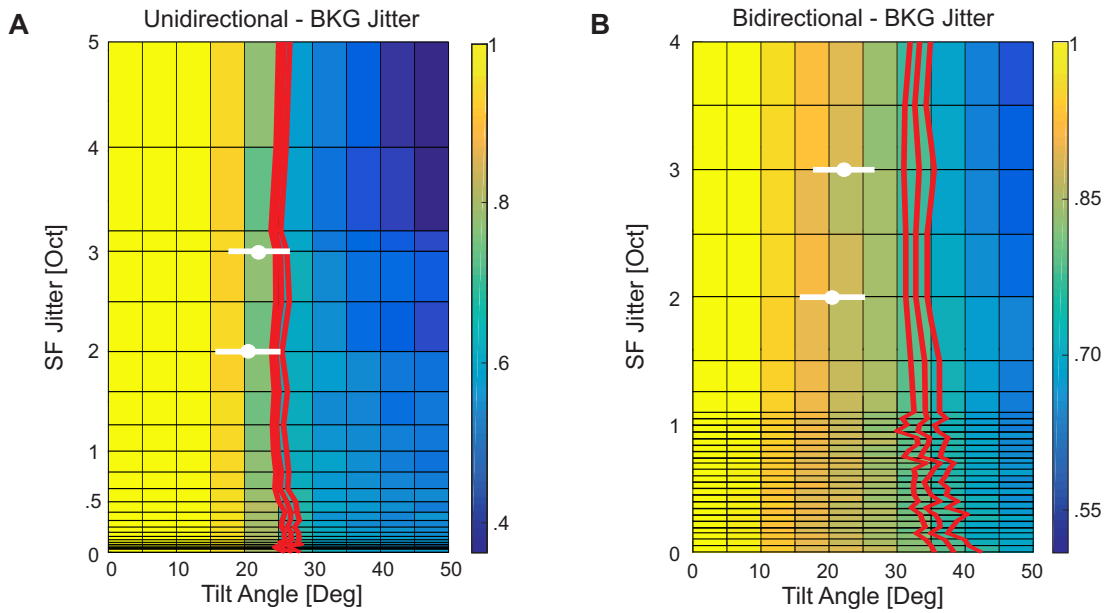

Figure 4.43: Proportion correct matrix for detection performance of contours in a display with spatial frequency jitter on background elements and tilt angle jitters causing deviations from perfect alignment (*Fixed Scaling Model*). **A**: Performance results yielded by a unidirectional model; **B**: Performance results yielded by a bidirectional model. Red lines are contour lines at 0.729, 0.75 and 0.776 proportion correct performance (i.e.: 0.75 proportion correct and upper and lower bounds of the expected error given by binomial statistics). White circles represent mean tilt angle threshold for humans to obtain a 75% correct performance at 2 and 3 octave jitters in spatial frequency (Persike and Meinhardt, 2015b), horizontal white lines intersecting the white circles represent standard deviation.

In this condition, contour detection performance was not (greatly) affected by the SF preference of background neural units because C to C connections were unaffected

(see figure 4.18 panel **B** and its associated discussion for details). Since excitatory C to C connections remained unaffected, the mean estimator (eq. 4.17) correctly identified a contour stimulus (as long as the contour was aligned to a certain degree, thus delivering excitatory C to C connections) regardless of the level of SF jitter.

Variable Scaling

This model did not perform as required. As can be seen by the red lines in both panels of figure 4.44 the tilt angle required for the model to perform at 75% correct was slightly reduced as jitter levels increased. This can be explained by the properties of lateral connections. Although the length scale of C to C and C to B lateral connections remained at λ_{ex_0}, the varying length scales of B to C and B to B lateral connections lead to a slight decrease in performance. As results showed for the *Fixed Scaling Model*, when lateral connections are homogeneous (and $\lambda_{ex_{ij}} = \lambda_{ex_0}$ for all ij) the results observed by Persike and Meinhardt (2015b) are reproduced.

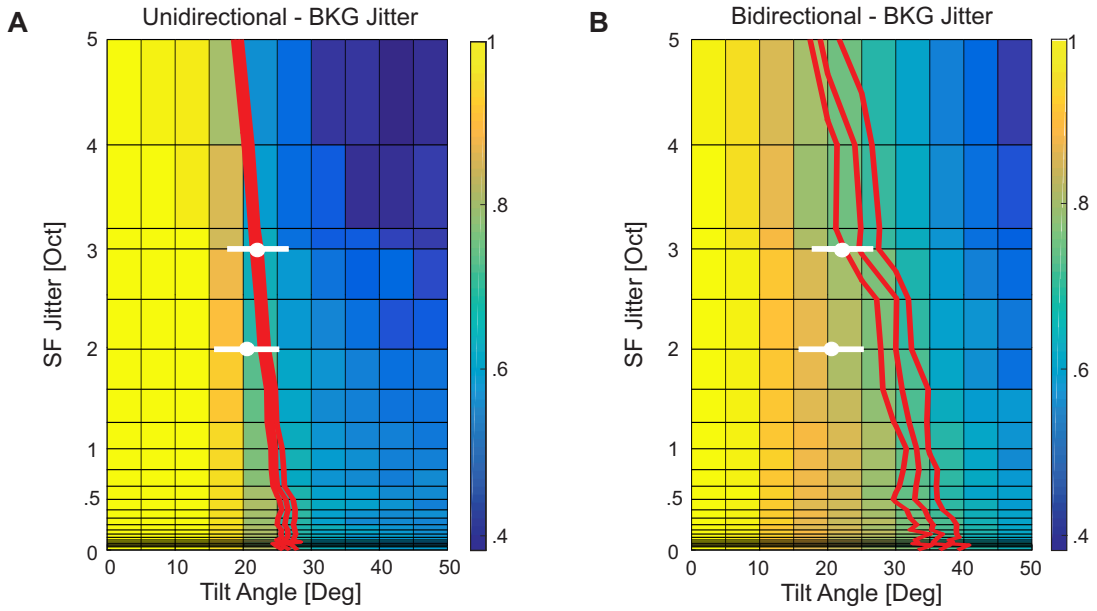

Figure 4.44: All conventions as in figure 4.43 (*Variable Scaling Model*).

4.6.2 Secondary Validation: Reproduction of Established Contour Integration Phenomena

In section 4.6.1, the model was validated with a number of psychophysical experiments performed by Persike and Meinhardt (2015a,b); Persike et al. (2009). The model was able to reproduce most of the results reported in the mentioned papers. As a secondary validation, the model was presented with a number of test scenarios inspired by well known psychophysical phenomena. In the following subsections, results for these test scenarios are discussed.

4.6.2.1 Reduction of Contour Length

In psychophysical experiments contour length, in terms of contour elements, has been shown to increase performance with an increasing number of elements (Braun, 1999; Li and Gilbert, 2002). It is expected that the model will also show higher performance levels with increases in the length of a contour (in terms of contour elements). The longer a contour is, the higher the activity of neural units representing contour elements will be, because of recurrent interactions.

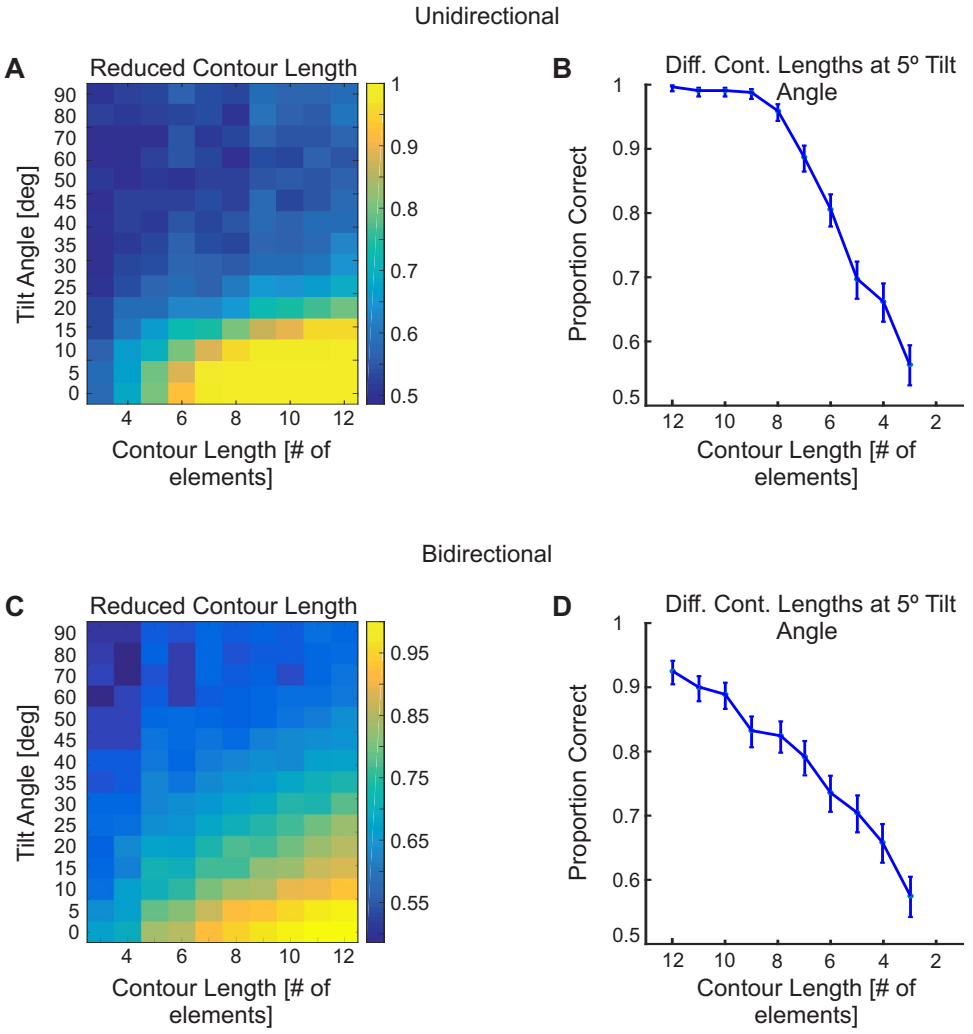

Figure 4.45: **A** and **C** correct contour detection performance matrix for tilt angle deviations from perfect alignment and different contour lengths; **B** and **D** Performance of contour integration with different contour lengths at a tilt angle of 5°. The vertical bars indicate the expected error given by binomial statistics, the round markers indicate average proportion correct contour detection performance for the given contour length at 5°. Panels **A** and **B** correspond to the unidirectional model, **C** and **D** to the bidirectional model.

Starting from a ceiling performance with 12 elements, at a tilt angle of 5°, the number of elements in contours were reduced. As can be seen in figure 4.45 performance decreases as contour length decreases. This is expected as: 1. the neural excitation

of units representing contour elements is higher if a contour is long rather than short (because of recurrent interactions); and 2. at short contour lengths (e.g.: 2-3 elements), there will be similar contour structures in both target and distractor stimuli.

4.6.2.2 Contours with Parallel Edge Configurations

In contour integration literature, two types of contours are usually studied: "snakes" and "ladders" (Bex et al., 2001; May and Hess, 2008; Vancleef and Wagemans, 2013). Snakes are the types of contours which have been used as stimuli for all previously reported results. The individual contour elements in these contours are aligned to the contour path. In ladder contours, the individual contour elements are perpendicular to the contour path (see figure 4.46 for examples). Field et al. (1993) and Ledgeway et al. (2005) have found that although contours are visible in ladder configurations, performance is not as high as it is in snake configurations.

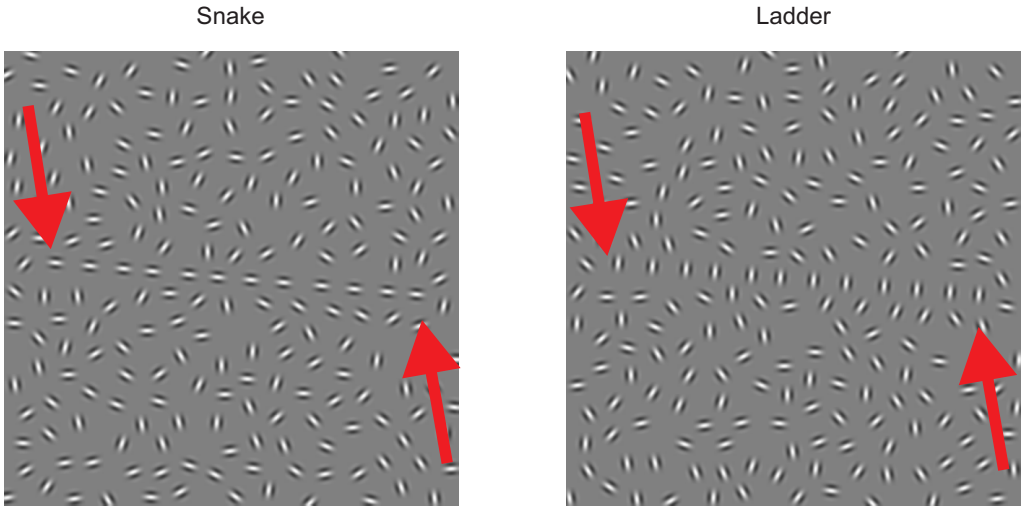

Figure 4.46: Examples of a snake and a ladder contour stimuli. Stimuli generated by Persike.

The performance found with ladder configurations is indeed not as high as that found for snakes. When snake contours are perfectly aligned to the contour's path, contour detection performance is 100%. Whereas, when elements in ladder contours are perfectly orthogonal to the contour's path, contour detection performance is approximately 60%. In ladder contours, detection performance is higher than chance (50% performance in a 2AFC). However, performance in ladder configurations does not reach the level which it does with snake configurations, because the angular part of the association field (equation 4.10) was optimized for snake configurations. The parameter σ_α in $W^{ex,ang}$ (eq. 4.10) is the length scale on which connection strengths decrease with distance from a co-circular edge configuration. In a perfect ladder configuration edge elements are oriented parallel to one another, thus, there is no optimal choice for the σ_α parameter.

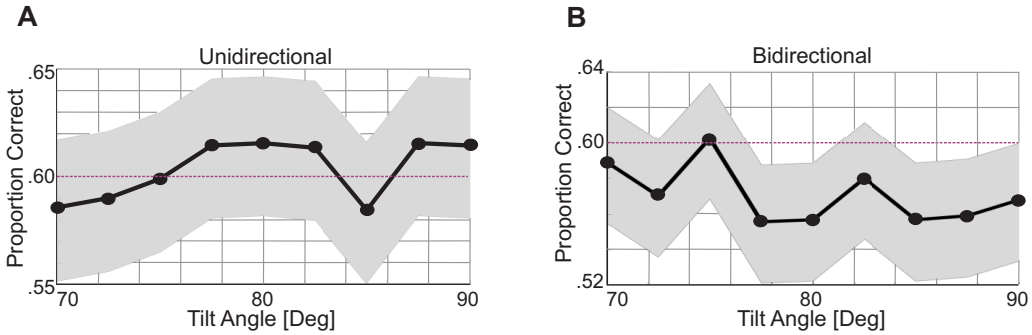

Figure 4.47: Model data for correct contour performance detection of ladder configurations. **A**: Unidirectional model; **B**: Bidirectional model. The shaded areas indicate the expected error given by binomial statistics, the round markers indicate average proportion correct contour detection performance for the given tilt angle. The dashed-magenta line marks performance at 60% for comparison purposes.

4.6.3 Predictions

After having validated the model through the reproduction of several psychophysical results, the model was presented with novel stimuli in order to characterize its behaviour. Knowing the way in which the different mechanisms were implemented (and how these mechanisms interact), allows for qualitative predictions to be made on the expected response of the model to new stimuli. It is interesting, first, to confirm whether the expectations will be corroborated or not. And second, to use the model as a test bed for new stimuli to be employed in future psychophysical experiments. Corroborating expectation on how the model should react also corroborates the modeler's understanding of the model, and testing stimuli on a model can help in the design and fine tuning of said stimuli.

In the following subsections, results of simulations done with novel stimuli will be reported.

4.6.3.1 Jitter on Contour Elements

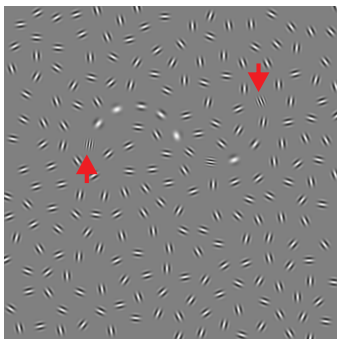

Figure 4.48: Target stimulus with **SF jitter on contour elements**

The following prediction developed while working with this model: when spatial frequency jitter is applied to aligned contour elements (while keeping the spatial frequency of background elements homogeneous), an increase in contour detection performance should be expected as jitter leaves increase. This prediction was based on known relationships between neural units of contour elements, and neural units of background elements. By applying a spatial frequency jitter on contour elements, the magnitude of neural activity exhibited by neural populations corresponding to said elements should slightly decrease. This is expected since there will be heterogeneity in the spatial frequencies of this group of elements. This heterogeneity will lead (on average) w_{ij}^{f} (eq. 4.9) to decrease for connections in which a contour unit is the presynaptic unit. This implies that the magnitude of contour-to-contour connections will decrease, and also that of contour-to-background connections. This effect, overall, will lead to a gain in the neural activities of populations stimulated by contour elements. An increase in these neural units' activity is expected since they will receive a lower amount of input from background neural populations (which tends to be mainly inhibitory).

Fixed Scaling

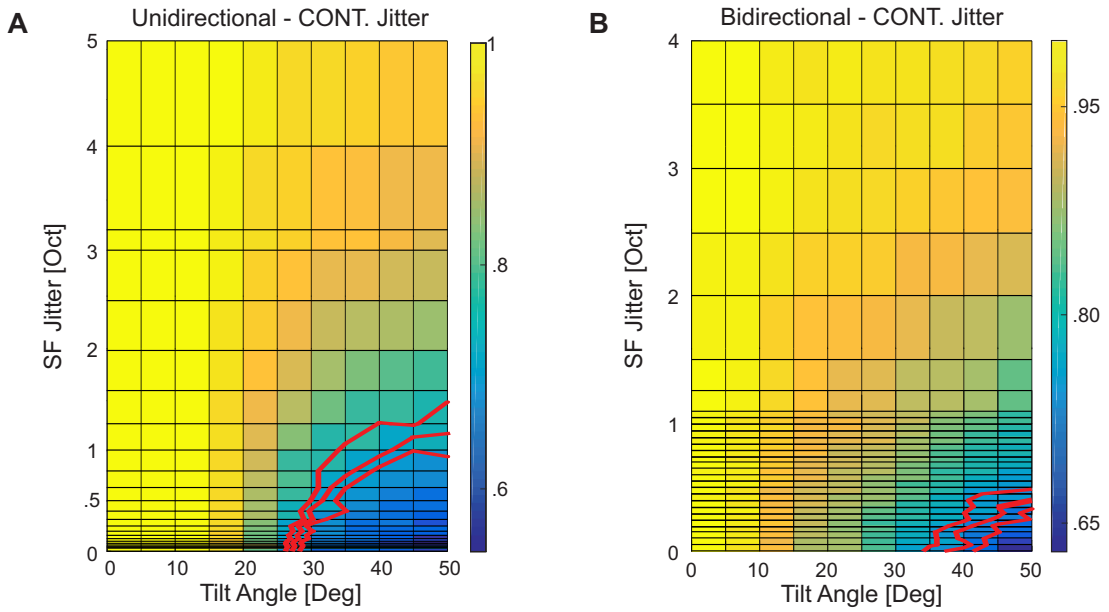

Figure 4.49: Proportion correct matrix for detection performance of contours in a display with spatial frequency jitter on contour elements and tilt angle jitters causing deviations from perfect alignment (*Fixed Scaling Model*). **A**: Performance results yielded by a unidirectional model; **B**: Performance results yielded by a bidirectional model. Red lines are contour lines at 0.729, 0.75 and 0.776 proportion correct performance (i.e.: 0.75 proportion correct and upper and lower bounds of the expected error given by binomial statistics).

The model yielded the results expected, and produced a qualitative reproduction of unpublished experimental results. As seen in figure 4.49, the ability of the model to detect contours became more resilient to deviations from perfect alignment as SF jitter

increased. In figure 4.50, three psychometric curves at fixed levels of contour jitter are shown. When compared with the 0 jitter case, the 2 and 3 octave jitter cases barely drop in detection performance over increasing tilt angles.

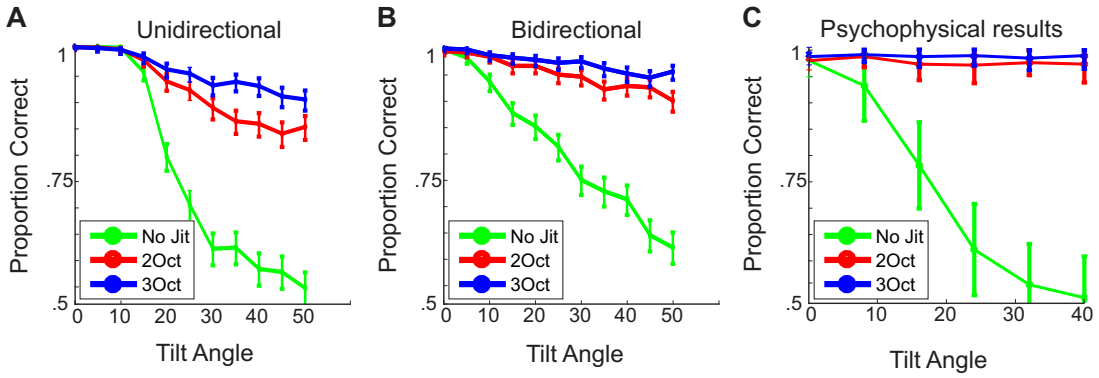

Figure 4.50: *Fixed Scaling Model* results at 0, 2, and 3 oct. of Contour Jitter. Detection performance of contours with spatial frequency jitter on contour elements and tilt angle jitter. Spatial frequency jitter shown are 0 octaves (green), 2 octaves (red), 3 octaves (blue). **A**: Unidirectional model; **B**: Bidirectional model; **C**: Unpublished psychophysics data provided by Malte Persike. In panels **A** and **B** vertical bars represent upper and lower bounds of the expected error given by binomial statistics, in panel **C** vertical lines represent standard deviations on the measured data.

Variable Scaling

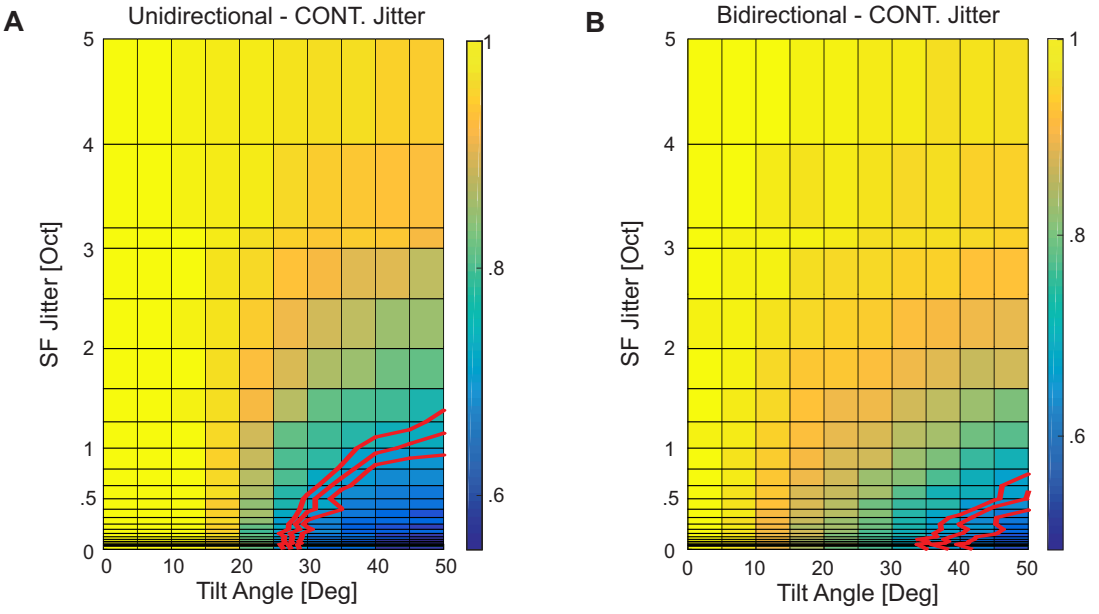

Figure 4.51: Proportion correct matrix for detection performance of contours in a display with spatial frequency jitter on contour elements and tilt angle jitter (*Variable Scaling Model*). Same conventions as in figure 4.49.

The *Variable Scaling Model* yielded the same pattern of results as the *Fixed Scaling Model* (as seen in figures 4.51 and 4.52). The variable range of lateral interactions in W^{ex} (for connections originating from contour units) did not have a large effect in contour detection performance. It seems that also in the *Variable Scaling Model*, results were mainly driven by the effects of W^{f} (eq. 4.9).

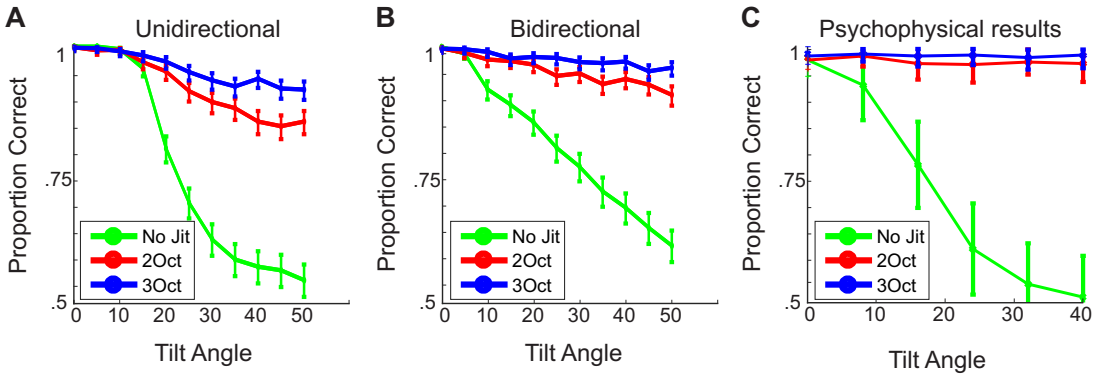

Figure 4.52: *Variable Scaling Model*. Same conventions as in figure 4.50.

4.6.3.2 Noise

The equation governing the dynamics of this model (eq. 4.1) allows for the introduction of noise into the system. Up to this point, all the results which have been presented have been modelled without noise. However, a noiseless system is not biologically realistic. There are several sources of noise in individual neurons and in the brain. In neurons, noise can occurs due to randomness in the cellular machinery, or due to nonlinear computations which occur within neurons (Faisal et al., 2008). In the brain there is background activity which is caused by random fluctuations in neurons' membrane potentials (Stern et al., 1997). There is often random exocytosis of vesicles which contain neurotransmitters, at times these bind to the postsynaptic membrane of other neurons. This process causes random fluctuations in the membrane potential of neurons, and it can even lead to the firing of action potentials (Fatt and Katz, 1952).

Efferent dynamic synaptic noise was simulated by means of equation 4.1 ($\eta_i = \sigma^{Noise}\xi\sqrt{T}$). Hence, the noise for each neural population modelled was unique, and it varied over time. With increasing levels of σ^{Noise} the information about the contour becomes less evident in the neural activity, thus, performance is expected to decrease as σ^{Noise} increases. In figure 4.53 the effects of different levels of noise in the **ORI** condition are shown. A comparison between contour detection performance by the model, stimulated with different levels of noise at different tilt angles, is made with human contour detection performance.

As can be seen in figure 4.53 panel **A**, for the unidirectional model, at approximately a level of $\sigma^{Noise} = 0.4$ the model delivers the closest results to the human data. At low levels of tilt angle jitter there is no difference between the human fit and the results of the model yielded by different levels of noise. This is because both the human fit and most levels of noise tested (except for $\sigma^{Noise} = 2$ and $\sigma^{Noise} = 4$) are at a ceiling

performance of 100% for low tilt angle jitter levels. However, in the mid range of angle jitter (between approximately 10 and 20°) the introduction of noise had a strong effect in performance, and brought the model results closer to the human data. This is especially seen for a model with $\sigma^{Noise} = 0.4$. This level of noise delivered almost an exact match in performance to the human fit, from 0 to 25° of tilt angle jitter. A more detailed comparison of this condition can be seen in figure 4.54.

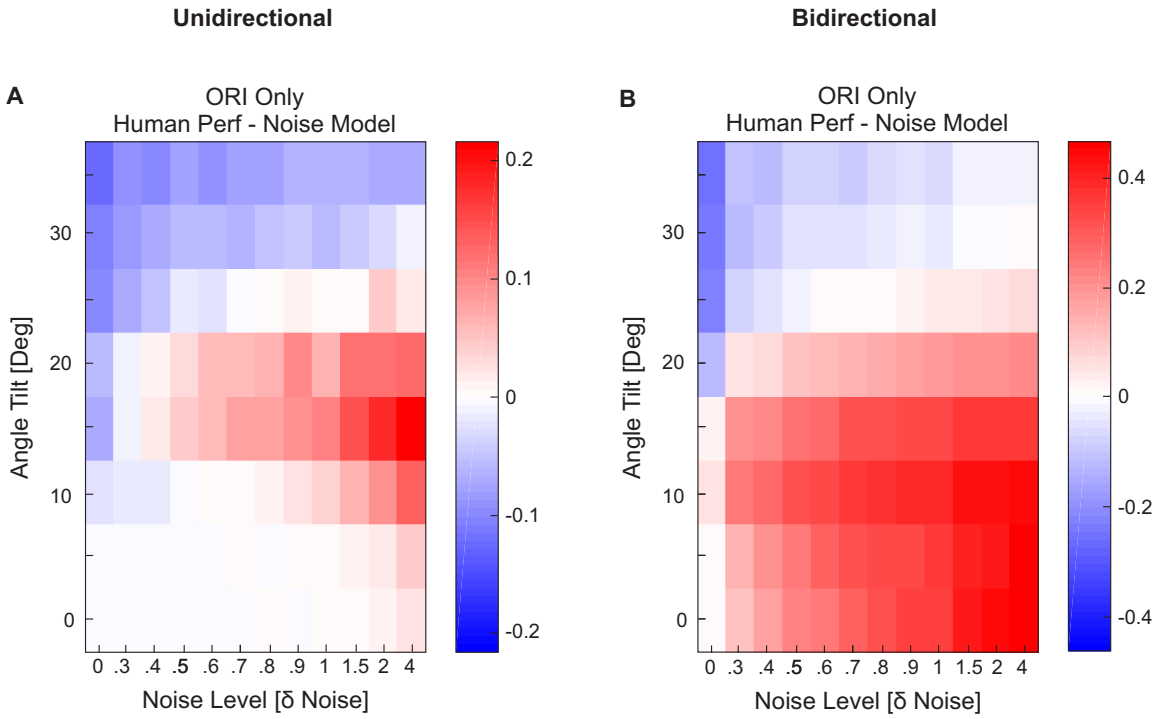

Figure 4.53: Proportion correct difference between human fit data and models stimulated with varying levels of noise in the **ORI** condition. **A**: Unidirectional model; **B**: Bidirectional model. Negative values indicate a higher performance of the model than psychophysical data.

Both, figure 4.53 and figure 4.54 show that for the **ORI** condition there were large differences in how the noise affected the model's performance. As discussed above, for the unidirectional model a level of $\sigma^{Noise} = 0.4$ led to a good match between the model and the human data. However, for the bidirectional model most levels of noise led to a worse performance of the model before approximately 15-20° of tilt angle, and a better performance thereafter. In the bidirectional condition, the psychometric curves yielded by the model did not have the typical "S" shape observed in psychometric functions. The decrease in performance as tilt angle deviations from the contour path increased led to a more monotonic drop in performance than in the unidirectional condition. Because of this, it is harder to evaluate the goodness of fit of these models to the human data (as the human data follows the typical "S" shape of psychometric curves). Furthermore, one of the critical requirements for a parameter set to be considered a good candidate to reproduce human data was the following: at all tilt angles tested, the the upper error bound for the model's performance should be higher, or equal, than the lower error bound for the human data. At 10° of tilt angle the chosen parameters for

fulfilled this requirement, but not by a large margin. The addition of noise which lead to a good performance match for high tilt angles (i.e.: $\geqslant 30°$), also lead to the model having a bad match (which did not fulfill the requirement stated) at low tilt angles (i.e.: $> 20°$). In summary, in the bidirectional model (with the chosen parameter set) a reduction in contour detection performance is clearly seen, however, a better fit for the **ORI** condition is not found when introducing noise.

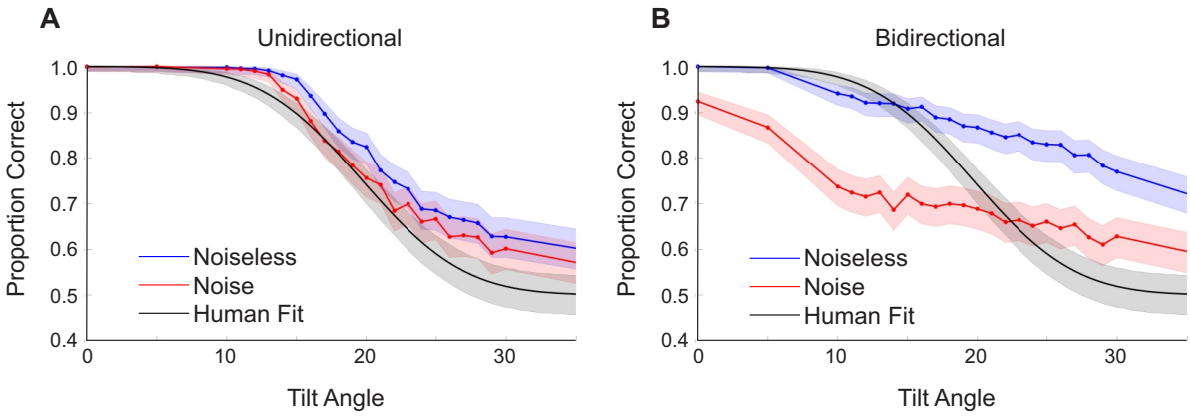

Figure 4.54: **Contour detection performance with and without noise**. Black curves: human fit, data obtained from Persike and Meinhardt (2015b). Blue curves: Noiseless models. Red curves: models with a noise level of $\sigma^{Noise} = 0.4$. Panel **A**: Unidirectional model; panel **B**: bidirectional model. Shaded areas represent confidence regions predicted by binomial statistics.

4.6.3.3 Contours with Varying Global Path Angles

So far, the model has always been presented with contours with a global path angle of 30°. These contours could take two general shapes, an O-shape or an S-shape (see figure 4.36 for examples). However, it is also possible to define contours with different path angles, and different shapes. Contours may have one point of inflection (e.g.: S-shaped contours), no inflection point (i.e.: straight contours), or a variety of different shapes.

Contour integration performance varies with different shapes and global path angles. Generally, the straighter a contour the more visible it is (Hess et al., 2001). In this section, results are reported for a set of stimuli with global path angles which could either be 0, 5, 10, 15, 20, or 25°. For simplicity purposes, these contours will be refereed to as 0, 5, 10, 15, 20, or 25° contours, respectively. The contours were of the O- and S-shape kind, and also varied in the number of elements which made up the contour. The possible number of contour elements were either 8, 10 or 12. For an example of the stimuli see figures 4.55 and 4.56.

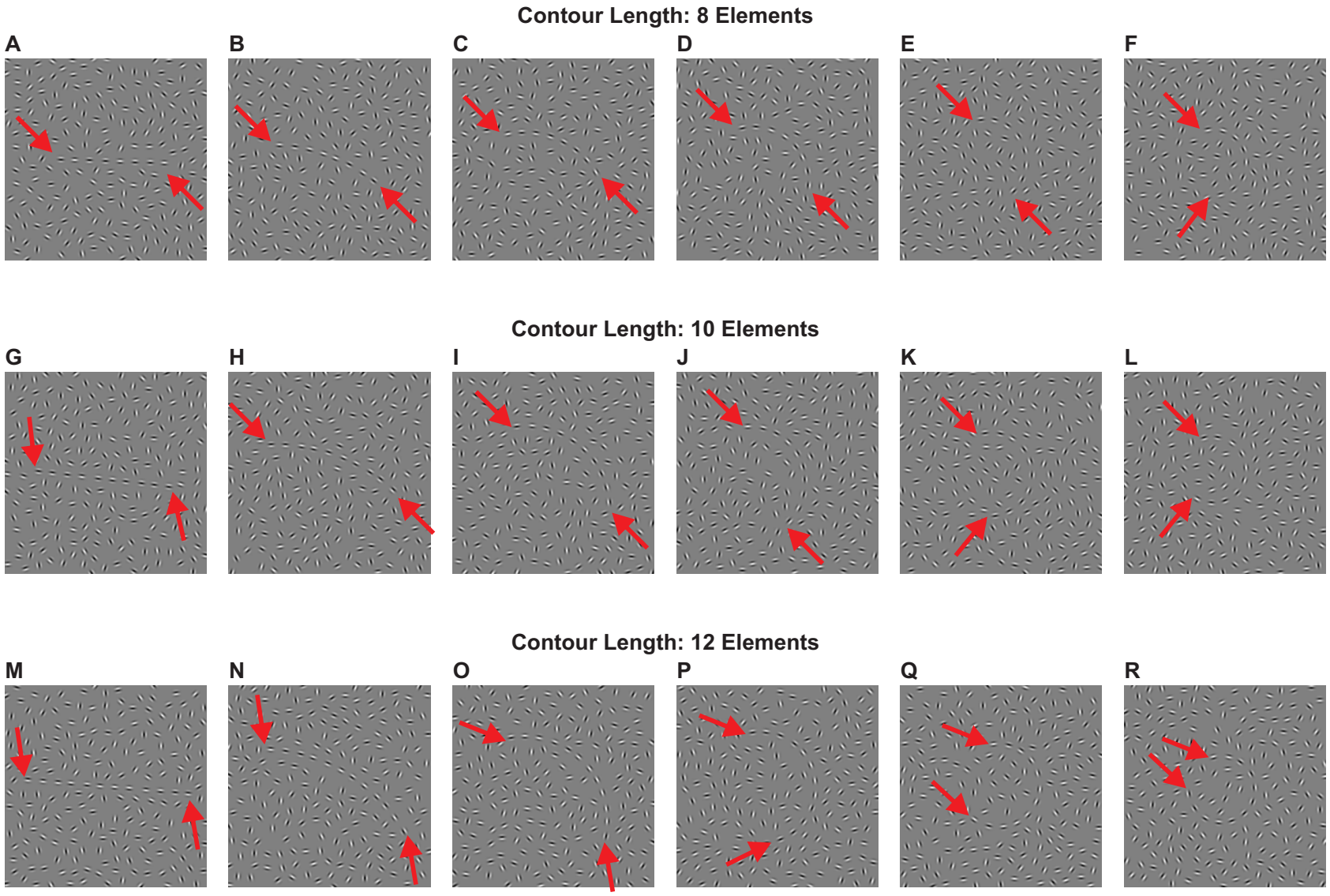

Figure 4.55: O-shaped stimuli with the indicated number of elements and global path angles ranging from 0 to 25°.

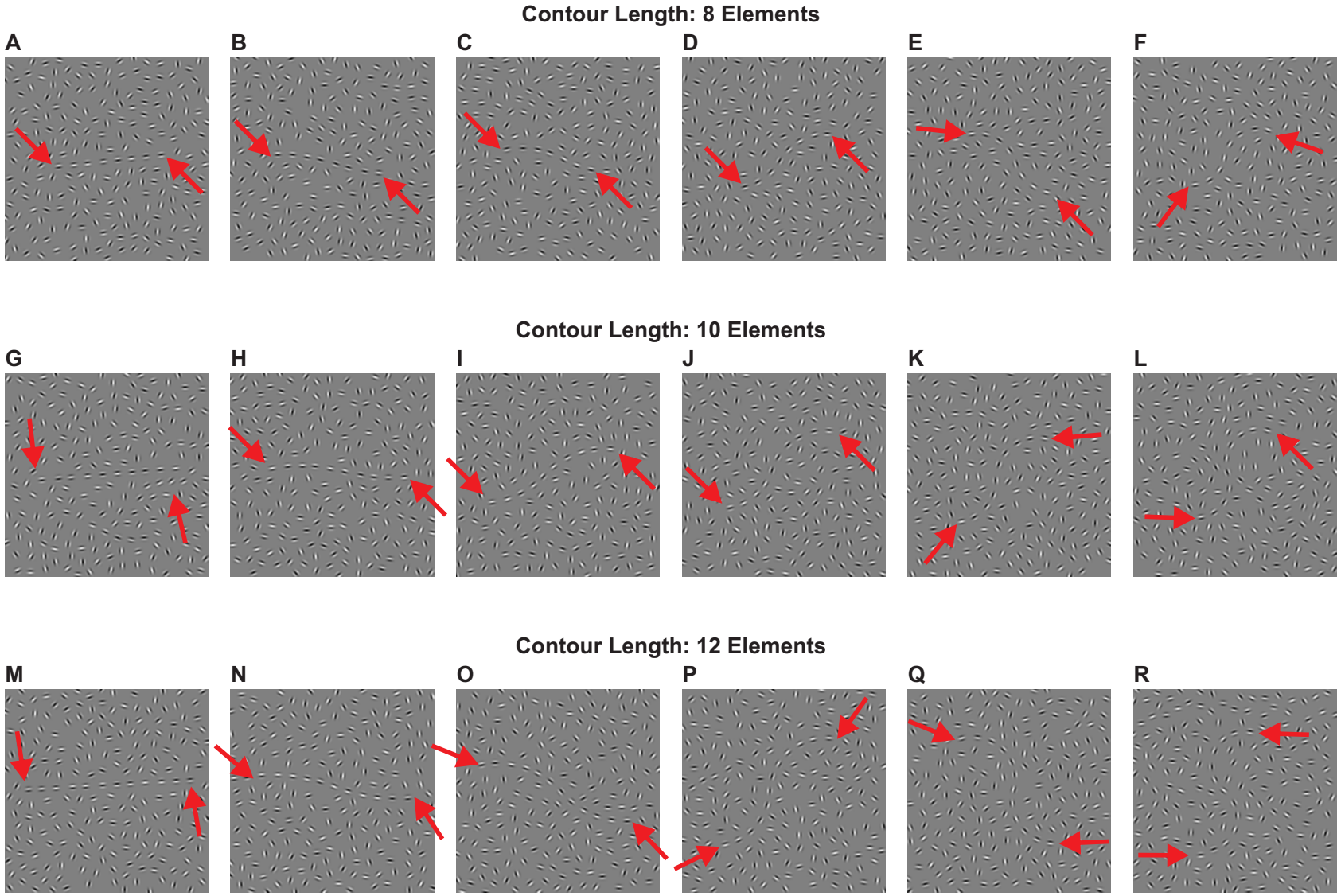

Figure 4.56: S-shaped stimuli with the indicated number of elements and global path angles ranging from 0 to 25°.

Figures 4.57 and 4.58 show the 75% correct contour detection performance tilt angle threshold (from now on referred to as tilt angle threshold) for O- and S-shaped contours, respectively. Each of the panels represent stimuli with a different number of contour elements (as indicated by the graphs' titles), in all panels data for all six possible path angles are shown in red. In black, a 75% tilt angle threshold for 30° contours is shown. The data corresponding to this type of contour is singled out as it was with this contour type that all previous results reported in this chapter were obtained.

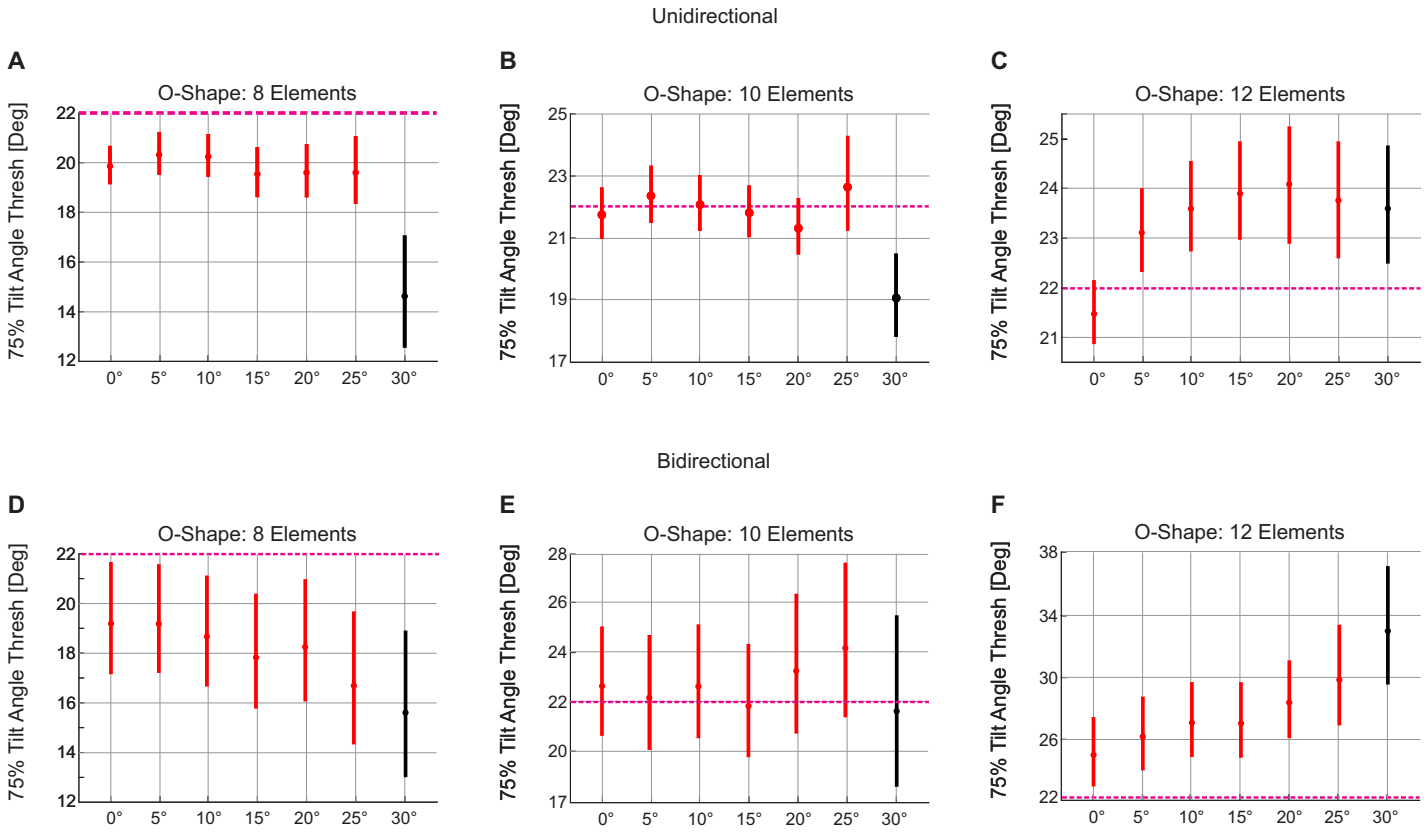

Figure 4.57: 75% tilt angle threshold for O-shape contour detection performance. Data in red corresponds to contours with global path angles of 0, 5, 10, 15, 20, 25°, in black to 30°. The 30° global path angle contour is singles out as this was the type of stimulus used for all other simulations presented in this chapter. Magenta lines mark the 22° tilt angle, facilitating results comparisons between the different panels. Upper three panels correspond to the unidirectional model, lower three panels correspond to the bidirectional model. Each of the panels correspond to the contour shape and length indicated by the title. Vertical lines indicate confidence regions predicted by binomial statistics.

Results for the unidirectional model, stimulated with contour lengths of 8 and 10 elements (panels **A** and **B**), show a pattern of better performance for 0 to 25° contours (inclusive), than for 30° contours. That is, with 0 to 25° contours the model was more resilient to individual element deviations from a perfect orientation to the path angle. These are similar results to what would be expected from human observers, as

these contours were straighter than the 30° contours (human subjects perform better with straighter contours (Hess et al., 2001)). For the bidirectional model, stimulated with contour lengths of 8 and 10 elements (panels **D** and **E**), a similar pattern can be observed. However, due to the large error margins it is hard to draw strong conclusions (especially for S-shaped contours with 10 elements - figure 4.58 panel **E**).

When the model was presented with 12 element contours results show a different and interesting result pattern. Panels **C** and **F** (in both figures) indicate that for contours of this length, 30° contours are as resilient, or even more resilient, to deviations of individual elements from the global path angle. This is most clearly seen in figure 4.57, panel **F**. Thus, this condition (O-shaped, 12 elements, bidirectional model) shall be taken to perform further analysis.

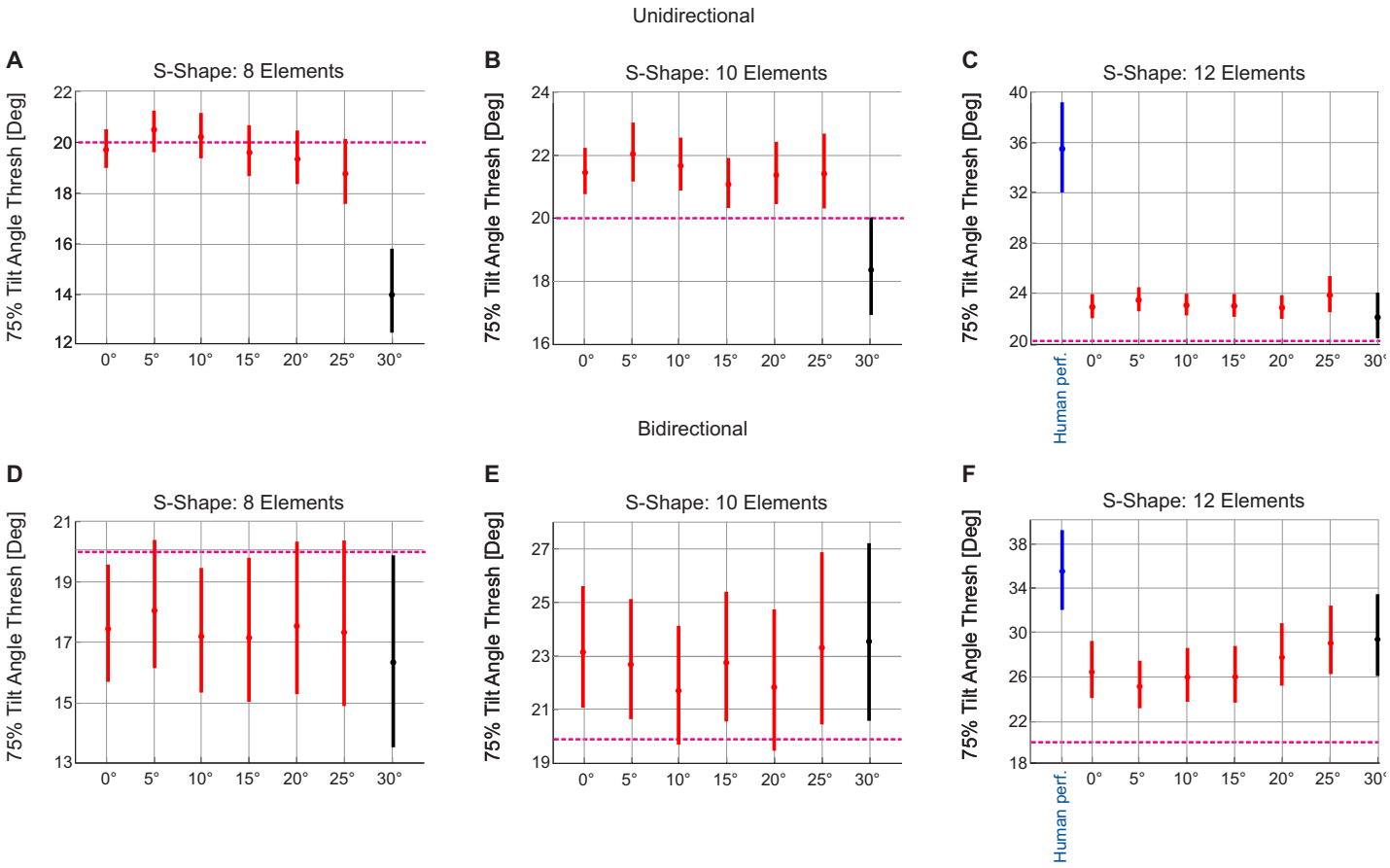

Figure 4.58: 75% tilt angle threshold for S-shape contour detection performance. Data displayed in blue in panels **D** and **F** correspond to unpublished psychophysics data. Remaining conventions are the same as in figure 4.57.

Due to the design of the association field employed, it would be expected that units stimulated by neighbouring elements which fall on the same line, and share the same orientation, would excited each other the most. If this was the case, then 0° contours should have the highest resistance to deviations of individual elements to the contour's path angle. However, in this case (O-shaped, 12 elements, and particularly the bidirectional model), 30° contours delivered the most resistance to these deviations. This claim is made since the data shows that 30° contours yielded the highest tilt angle

threshold. Figure 4.59 shows the mean neural activity of contour units, for contours with path angles ranging from 0° to 25°. Panel **A** shows that at a 0° tilt angle, contours with a lower global path angle led to the greatest mean activity of neural units (stimulated by contour elements). However, at tilt angles between 20° and 40° this pattern starts to change. In this range, the mean activities corresponding to contours with larger global path angles become higher. This change occurs because of an interplay between the global path angle of the contour, and the local orientation of elements.

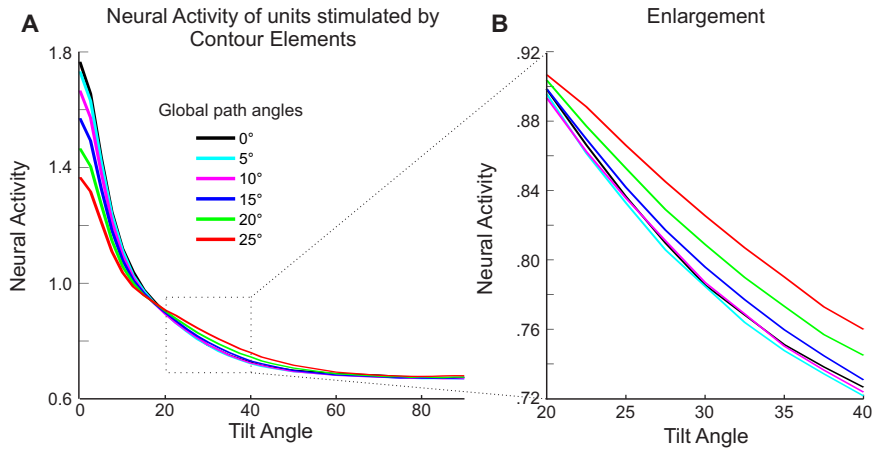

Figure 4.59: Mean activity of neural units stimulated by contour elements in a display with target contours of length 12. Data for contours with global path angles of 0, 5, 10, 15, 20 and 25° is shown, as labeled. Panel **B** is an enlargement of the area marked in panel **A**. The data corresponds to O-shaped stimuli and a bidirectional model.

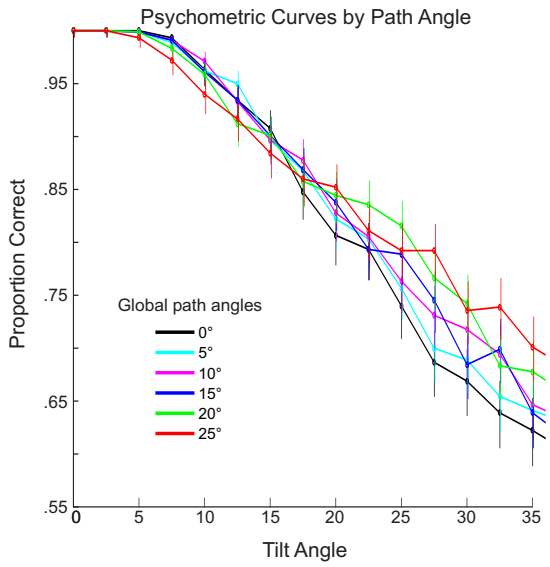

Figure 4.60: Correct contour detection performance over tilt angles. Data for contours with global path angles of 0, 5, 10 15, 20 and 25° is shown, as labeled. The data corresponds to 12 elements O-shaped stimuli and a bidirectional model.

If contour detection performance is obtained for contours with different global path angles, the order in which performance curves are arranged at low tilt angles and at high tilt angles will differ. Figure 4.60 shows performance for contours with different global path angles. It is evident that performance is highest for contours with a global path angle close to 0° when tilt angles range between 5° and 10°, but that this changes at higher tilt angles.

4.6.3.4 Singleton

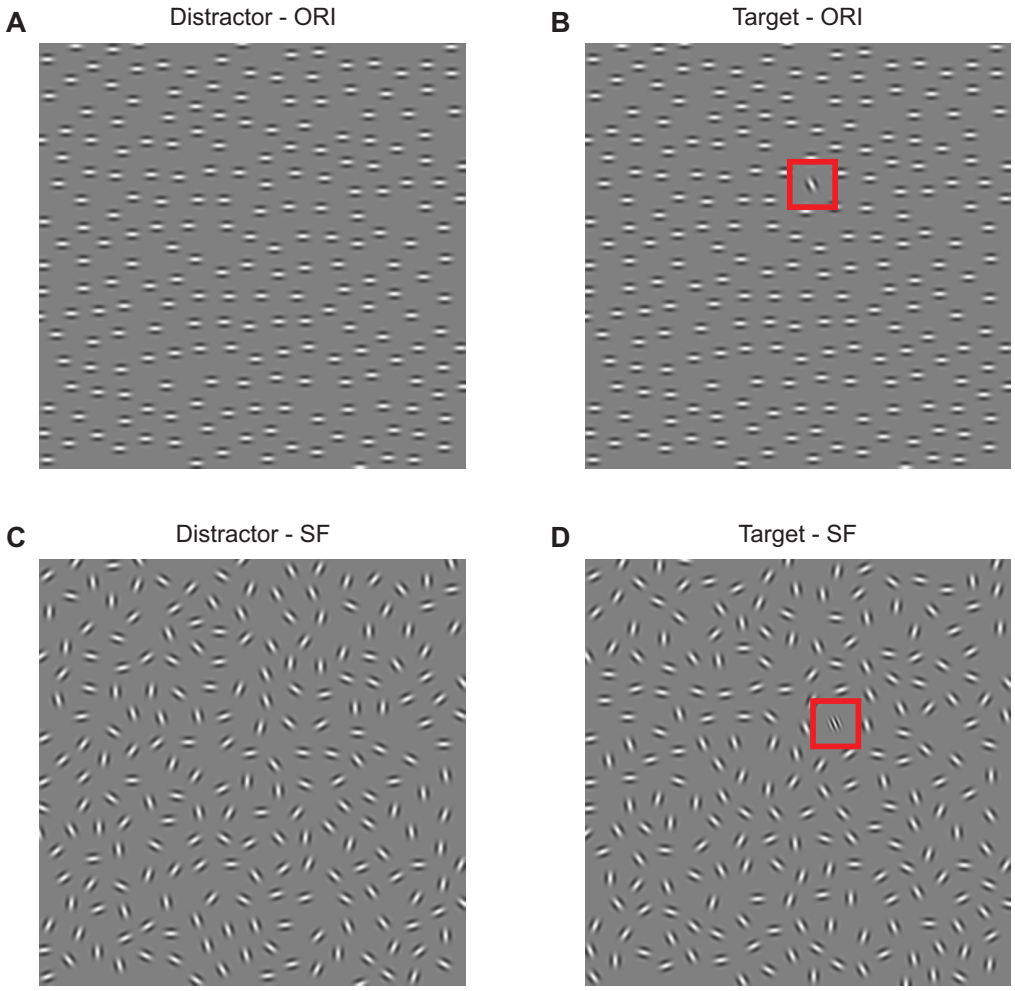

Figure 4.61: Example stimuli for the singleton conditions. **A**: distractor for the **ORI** condition; **B**: target for the **ORI** condition; **C**: distractor for the **SF** condition; **D**: target for the **SF** condition. Red square placed around the target element in panels **B** and **D** to help the reader localize the singleton.

An interesting condition in which to analyze the behaviour of the model is that of a singleton (see figure 4.61 for samples of all stimuli mentioned in this section). In the **ORI** condition, by giving all gabor elements the same orientation it is possible to evaluate the response of the model to the tilting of one element (stimuli examples: figure 4.61, **A** and **B**; model results: figure 4.62, **A** and **C**). By giving a random orientation

to all elements, and shifting the spatial frequency of one of the elements with respect to that of all others, the individual effects of spatial frequency differences can be analyzed (stimuli examples: figure 4.61, **C** and **D**; model results: *Fixed Scaling Model* - figure 4.62, **B** and **D**; *Variable Scaling Model* - figure 4.63). The combination of spatial frequency differences and orientation differences of one element can be evaluated by tilting one element in a homogeneously aligned field, and giving this element a distinct spatial frequency in an otherwise SF homogeneous field of Gabors. However, with the current model configuration, problems arise when reading out the neural activity of this situation. Thus, it is not shown in this section. A further discussion of this topic ensues below.

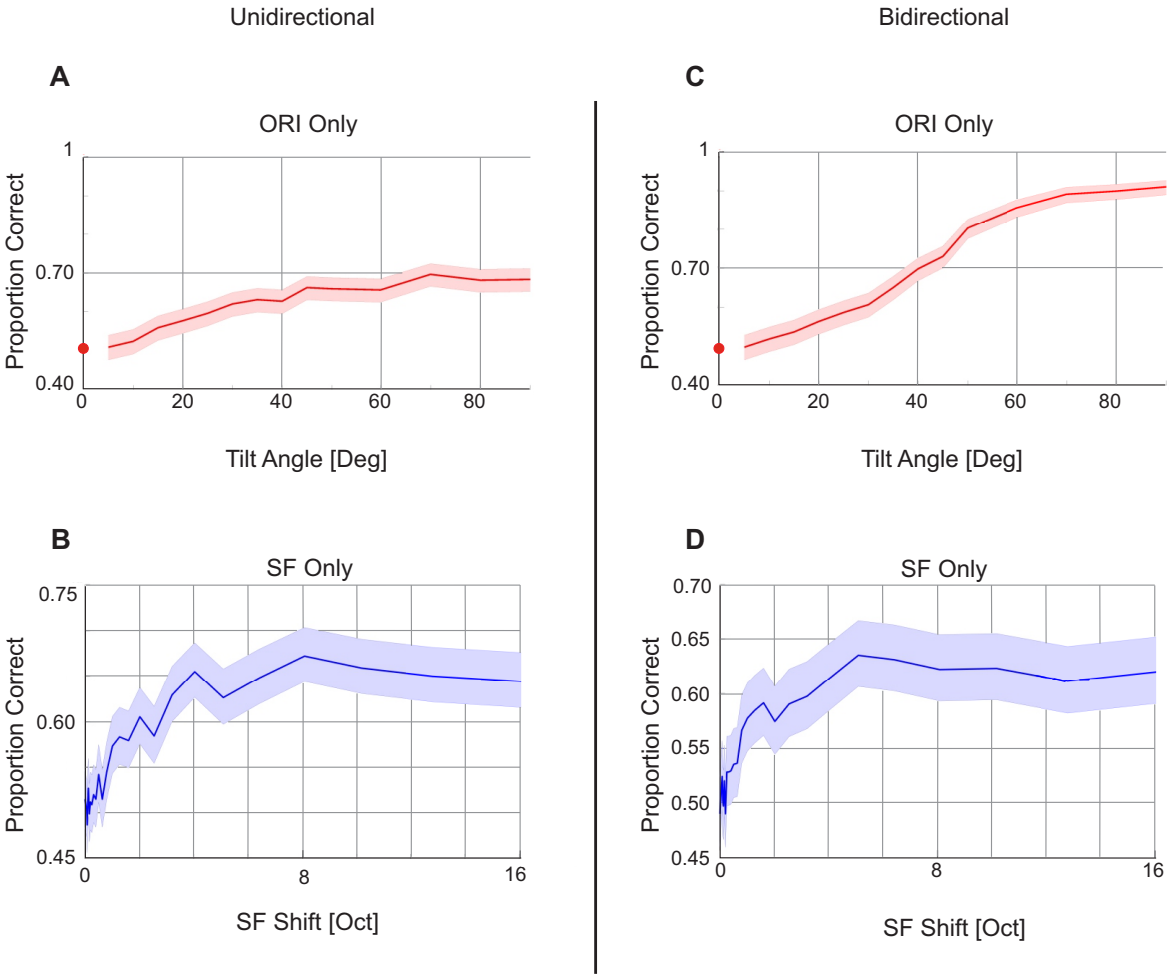

Figure 4.62: Detection performance for Singletons with varying properties. **A** and **C**: **ORI** condition, unidirectional and bidirectional model data respectively. **B** and **D**: **SF** condition, unidirectional and bidirectional model data respectively. Note that the sum estimator was used for the **SF**, and the inverse of the sum estimator was used for the **ORI** condition. *Fixed Scaling Model* data.

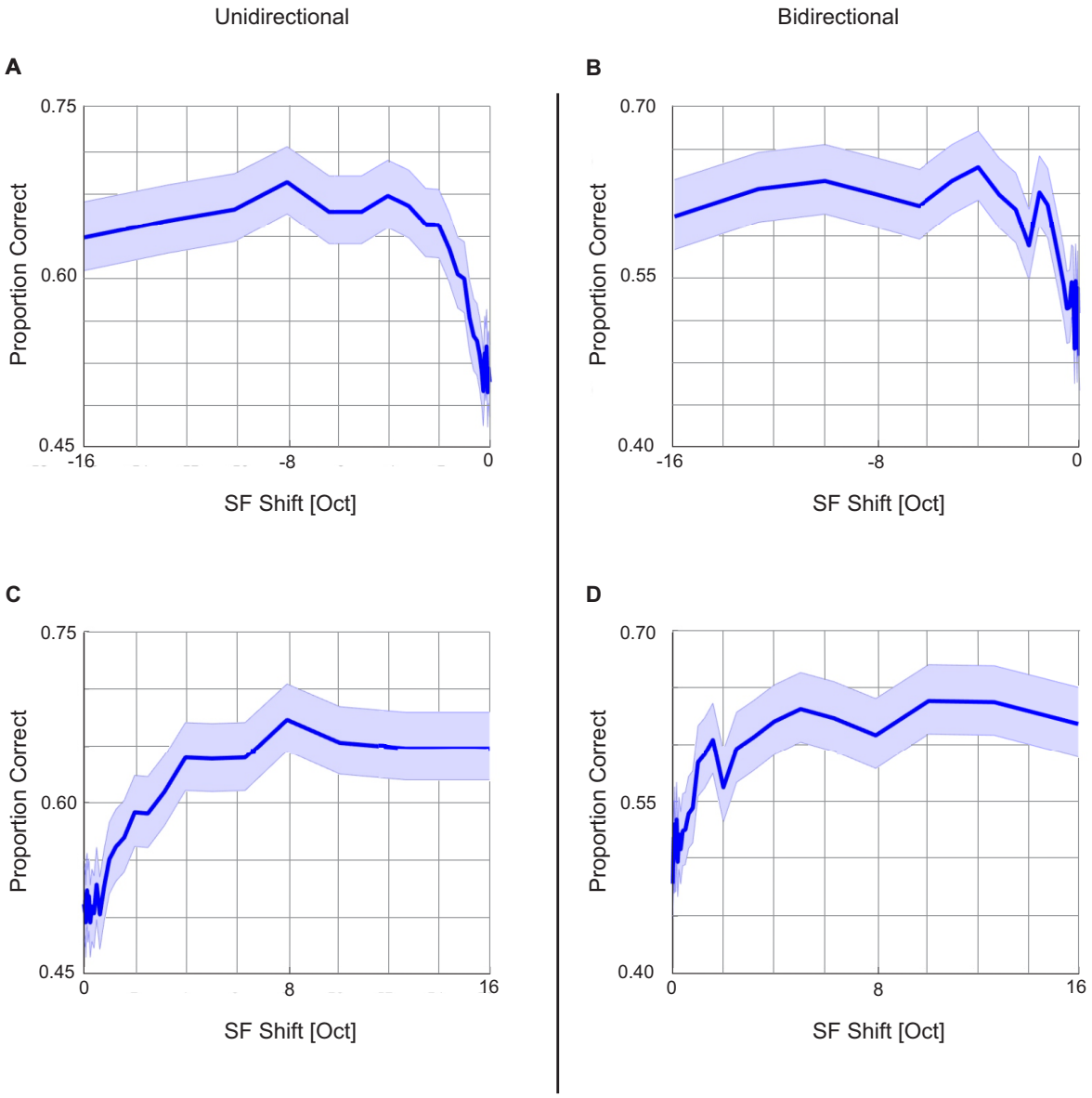

Figure 4.63: Detection performance for **SF** singletons in the *Variable Scaling Model*. **A** and **B**: $\mathbf{SF}_{\downarrow}$ condition; **C** and **D**: $\mathbf{SF}_{\uparrow}$ condition. Unidirectional and bidirectional models as marked.

It is important to note that in the **ORI** condition, in order to obtain psychometric curves which increase in proportion correct with increasing tilt angles, it was necessary to take the inverse of the sum estimator (see equation 4.16). This was required because the distractor, a field of gabor elements with the same orientations, lead to higher activities in the neural units than the target did. The target was composed of a field of Gabors in which all, elements except one, shared the same orientation. Since the distractor contains "better contour structures" than the target, it was expected that neural activities would be higher with distractor stimuli. It was also necessary to use ensembles with Gabors located at the exact same locations as target and distractor stimuli. This was required because with a high degrees of homogeneity, having a distinct number of elements in target and distractor ensembles can drastically bias the decision

of the estimator. Because neural activities arising from ensembles with almost equal properties were compared, it was also necessary to apply the sum estimator, rather than the mean estimator. This decision was taken because the subtle differences in overall neural activities arising from such homogeneous displays might be lost when taking an average.

Other peculiarities arise in the singleton **ORI** condition, due to the high degree of homogeneity between the target and distractor ensembles, and because the estimator was not defined precisely enough. In panels **A** and **C** of figure 4.62, a single point can be seen at 0° of tilt angle and at 0.5 proportion correct. This point was artificially placed there as it is the expected result under the described conditions. The estimator is not well defined for cases in which the model is presented with a target and distractor with a high degree of homogeneity. When the neural activity resulting from the target and distractor are the exact same, the estimator would yield 100% correct performance result. Such is the case because a "greater than" binary decision between two identical levels of neural activity is made by the estimator (two identical neural activity levels are compared because two stimuli which are identical in all respects were used to stimulate the model in this condition). However, if noise is introduced, over a large number of stimuli tested, proportion correct will approximate 0.5.

In the singleton **SF** condition, it was not necessary to employ the inverse of the sum estimator (as was the case in the singleton **ORI** condition). In the singleton **SF** condition, target stimuli lead to higher neural activities than distractor stimuli. In this condition, target stimuli contained random orientations for all elements **and** one element which was distinct in terms of spatial frequency. Distractor stimuli were composed of randomly orientations elements, with a homogeneous SF for all elements. It was not necessary to employ the inverse of the sum estimator, as was the case in the singleton **ORI** condition, because in both target and distractor stimuli, all elements had random orientations. Thus, (on average) neural units did not receive excitation. As such, reducing the magnitude of connection strengths between elements (by changing the spatial frequency of one element, hence influencing w_{ij}^{f} for all ij which involved the heterogeneous element - see eq. 4.9) led to higher neural activities, since connections were predominantly inhibitory.

In figure 4.63, results for the *Variable Scaling Model* are presented. These do not vary significantly in the $\mathbf{SF}_{\uparrow}$ and $\mathbf{SF}_{\downarrow}$ conditions, nor are these very different from the *Fixed Scaling Model* results. The performance increase were mainly dominated by the effects of a change in W^{f}, rather than by the change in reach of w_{ij}^{ex}'s lateral interactions. As was the case for the *Fixed Scaling Model*, the increase in performance peaks at about 8 oct. Δ SF, and stabilized thereafter.

The contradiction of having to use the inverse of the sum estimator in the **ORI** condition, and the regular sum estimator in the **SF** condition, led to the impossibility of properly reading out the combination of an **ORI** and **SF** condition. Thus, this situation was not tested.

For both single feature conditions the model was able to reliably detect a singleton. In the **ORI** condition, the singleton was detected by a decrease in neural activity, because in this condition the target was a "bad contour structure" (i.e.: the opposite of what the model was designed to detect). In the **SF** condition, the model was able to

detect the singleton with the same read-out mechanisms which was normally applied, as the target ensemble lead to a higher neural activity than the distractor ensemble.

4.7 Summary and Discussion

The results from the reported modeling experiments help provide a more comprehensive picture of feature integration by the visual system. With the central hypothesis that interactions of feature detectors are stronger if their preferred spatial frequencies are similar, rather than dissimilar, a structurally simple neuronal model of contour integration was devised. The model was based on an existing contour integration model introduced by Ernst et al. (2012). The main differences between the model introduced by Ernst et al. (2012) and the current one are that: **a.** the current model is a neuronal model rather than a probabilistic model; **b.** the current model contains a spatial frequency component to modulate the interaction strengths between feature detectors. Introducing this component allowed for the reproduction of experimental evidence showing spatial frequency to be a strong cue on which it is possible for the visual system to integrate contours (Persike and Meinhardt, 2015a,b; Persike et al., 2009).

The model was validated on a number of psychophysical results found by Persike and Meinhardt (2015a,b); Persike et al. (2009), and also by reproducing hallmark results in contour integration literature (e.g.: decrease performance with decreasing contour length, detection of ladder configurations, higher performance for closed contours, etc.). Most of these results could be reproduced by the model. Notably, in many cases the model outperformed humans. However, it was shown that a close match between human and the model's performance can be achieved by introducing noise into the system.

Predictions were made for how the model should behave when presenting it with number of distinct stimuli. The confirmation of these predictions suggest that although the model was dominated by non-linear dynamics, its behaviour could be well understood. Having a model which is capable of reproducing a large number of psychophysical results, and understanding its mechanisms, allows for the modeler to make testable hypothesis. Not only does this provide a great understanding of putative mechanisms which the visual system may employ while conducting contour integration, but it also allows for the development and testing of novel psychophysical stimuli/paradigms. Furthermore, having an algorithm which can outperform humans and is highly accurate in contour detection might prove to be useful in computer vision.

Although the model could reproduce many psychophysical results, and it has the potential to be useful for different purposes (e.g.: development of stimuli for psychophysical experiments; computer vision), it also has its shortcomings. A discussion of these ensues:

1. **Directionality:** Two variants of the model were presented, a unidirectional model and a bidirectional model. In the unidirectional variant two neural populations with their RF center at the same position were modelled for every edge element. These neural populations were assigned axons assumed to project in

opposite preferred directions. In the bidirectional variant only one neural population whose axons were assumed to project in polar opposite directions was modelled for every edge element. For details regarding these two variants of the model refer to section 4.4.2.

Typically, the bidirectional modelled performed better than the unidirectional model. This was due to the bidirectional model having a higher tolerance for tilt angle jitters than the unidirectional model. The reason for this higher tolerance is that the structure of the bidirectional model leads to greater activity in (orientation aligned) neural units, because recurrent interactions are more effective with this architecture. Although the bidirectional model typically outperformed the unidirectional model, the unidirectional model's results were usually a better match for psychophysical data. This presents a conflict. Biophysically, the bidirectional model is more plausible than the unidirectional model, since lateral connections spread isotropically between cells in the brain. However, it is also possible that the brain could perform contour integration in a manner akin to the unidirectional model. This would be the case if there is a refractory period in which the activity induced by the perception of aligned edge elements starts to spread in one direction, and then in another. As suggested by (Schinkel-Bielefeld, 2008), a contour has no direction per se. Thus, the activity of a cell which "sees" one of the contour elements could start to spread in one direction, and after a refractory period it could start to spread in the other. This would be, to some extent, similar to the unidirectional model in which activity only spreads in one direction, but there are two sets of neural units for every edge element which independently project in polar opposite directions.

2. **Fixed and Variable Scaling Models:** In order to account for one of the nuances in the results reported by Persike and Meinhardt (2015a) the fixed and the variable scaling model variants were introduced. In the *Fixed Scaling Model* the lateral reach of excitatory interactions was fixed, whereas in the *Variable Scaling Model* the reach of excitatory interactions depended on the spatial frequency preference of neural units. This assumption was based on experimental observations (Ernst et al., 2016; Polat and Sagi, 1993; Tolhurst and Barfield, 1978) showing that the lateral spatial reach of feature detectors varies according to their spatial frequency tuning. Although this assumption was biophysically plausible, it did not lead to the successful reproduction of one result reported by Persike and Meinhardt (2015a). When background neural units had a lower spatial frequency preference than contour neural units, the model's performance went below chance level. See section 4.6.1.2 for a detail discussion of this topic.

   By maintaining the reach of lateral interactions fixed, as in the *Fixed Scaling Model*, and by making the effects of feedforward input dependent on SF preferences of neural populations, it could be possible to reproduce all sets of results. With this new conception of the model, neural populations with a spatial frequency preference of f_0 would receive an external input of magnitude 1, as has been the case so far. However, if neural units have a spatial frequency preference other than f_0, their external input could be scaled so that it ranges between 0 and

1. If the function used to scale the feedforward input is skewed, then a spatial frequency preference of the same magnitude, but of opposite sign, would yield a different magnitude of feedforward input to these two neural units (i.e.: if one neural unit has an SF preference which differs from f_0 by x oct., and another neural unit has an sf preference which differs from f_0 by $-x$ oct., these two neural units would receive a different feedforward input). By modifying the feedforward input in this manner, neural activity would be amplified by different amounts, based on the spatial frequency preference of neural detectors.

   The decision to scale feedforward input based on the SF preference of neural units can be justified with experimental evidence. Contrast facilitation has been shown to be dependent on the spatial frequency of gratings (Enroth-Cugell and Robson, 1966; Solomon et al., 1999; Williams and Hess, 1998; Woods et al., 2002). Furthermore, Enroth-Cugell and Robson (1966) has shown contrast sensitivity functions to be negatively skewed. Thus, by choosing to scale the feedforward input with a negatively skewed function, with a peak at f_0, the model would deliver greater sensitivity to decreases in SF preferences (in relation to f_0) than to increases in spatial frequency preferences (also in relation to f_0).

3. **Choice of k_{mean} as estimator:** Several estimators were conceived as read-out mechanisms, however, k_{mean} was chosen to present performance results in this chapter. See section 4.5.1 for a detailed description of the read-out mechanisms.

   The estimators k_{sum}, k_{mean}, k_{square}, and k_{max}, rely on simple algorithms which can be performed very quickly and easily (computationally speaking). However, s^2, s^3, and s^4, rely on more complex computations, since they require access to the mean neural activity given by all the neural populations modelled. As such, biophysically k_{sum} ,k_{mean}, k_{square}, and k_{max} are more plausible than either s^2, s^3, or s^4, due to their simplicity (Li, 2014).

   It is assumed that when performing a visual search the brain relies on saliency maps. Similarly to the saliency maps created with this model, in the brain's "saliency maps" there are objects or areas in space which are assigned a higher saliency than others by the visual cortex. In order to complete a visual search these areas have to be identified. Thus, in the brain, the saliency maps created by the visual cortex need to be "read-out". The superior colliculus (SC) (Basso and Wurtz, 1997; Tehovnik et al., 2003) may be responsible for this process since it is responsible for directing gaze to areas of high saliency in a display, and has often been linked to behaviour based on attentional capture (Fischer and Breitmeyer, 1987; Goldberg and Wurtz, 1972; Müller et al., 2005). Thus, in the brain, the SC may perform the task of the readout mechanisms described in section 4.5.1.

4. **Singleton:** In section 4.6.3.4 the detection of singletons by the model was discussed. For the singleton which was defined on the basis of an orientation difference from all other elements in the display, the inverse of the sum estimator was employed as a read-out mechanism. However, for the case in which the singleton was defined on the basis of a spatial frequency difference from all other elements in the display, the regular sum estimator was employed. The employment of

two different estimators lead to the inability to properly read-out neural activity in a case in which the orientation and spatial frequency cases were combined. Furthermore, detection performance never reached 100% for either of these two cases.

Psychophysics experiments have shown that singletons are easily detectable and often cause strong attentional and oculomotor capture (Lamy and Egeth, 2003; Theeuwes, 1991; Theeuwes et al., 2003). As such, it would be expected that if the experiments conducted in section 4.6.3.4 were performed with humans, their detection performance should approximate 100% correct. In the modeling experiments performance did not reach 100% correct neither for the **ORI** condition, nor for the **SF** condition for the following reasons. **a.** In the **ORI** condition inhibitory interactions did not play a major role. Both target and distractor stimuli were virtually identical, and inhibitory interactions were implemented in an unspecified manner (meaning that inhibitory interactions from all neural units had the same strength, which did not depend on the characteristics of the stimulus). Thus, in this condition targets could only be identified on the basis of differences in neural activity which originated from excitatory interactions. In the target condition, after a certain tilt angle deviation, $w_{ij}^{ex,ang}$ reached its lower bound (for all ij pairs which included a connection to the element with a unique orientation). Thus, beyond that point, further deviations in orientation did not lead to a greater differences in neural activities in target and distractor trials.

b. In the **SF** condition, for both the *Fixed* and *Variable Scaling Model*, the increase in performance was mainly due to the effects of W^f. In these conditions excitatory interactions were (on average) negligible because there were no aligned elements, all the orientations were randomly assigned. Thus, in this condition the target was detected on the basis of effect caused by a reduction in inhibitory connection strengths, due to a decrease in w_{if}^f (for all ij pairs which included a connection to the element with a unique SF). Beyond approximately 8 Δ SF w_{ij}^f (for the mentioned connections) leveled out at its minimum. Thus, greater differences in the SF preference of the single unit and all others did not lead to changes in the connection strengths between neural units.

A possible solution for the issues discussed would be to implement inhibitory interactions in a different manner. In (Li, 1998) inhibitory interactions are modelled on the basis of orientation preference similarities/dissimilarities between feature detectors. If this approach were to be implemented in this model, and parameters are selected to optimize the detection of singletons, detection performance in the cases described could improve.

Future extensions of this model should address the issues presented above. Currently, work is being done on the issue discussed in the second point (i.e.: Fixed and Variable Scaling Models). Preliminary tests indicate scaling feedforward input in the manner described could be more effective than the method employed in the *Variable Scaling Model*, as it appears that with this method all results presented by Persike and Meinhardt (2015a) can be reproduced. However, further tests and a refine parameter search will be necessary before any conclusion can be drawn.

A mathematical analysis of the model would provide greater insight into the role that each of individual components in the model plays in generating the association field. This would possibly allow for more accurate predictions to be made, and perhaps for new testable hypothesis to be devised. A mathematical analysis would also provide provide a greater insight into proposed mechanisms for CI with orientation and spatial frequency cues, thus, providing a greater understanding of the functions of the brain.

As previously mentioned, this model could find applications in computer vision. If this model is to be employed in that field then a re-evaluation of the readout mechanisms should be done. For the computer vision field the selection of a readout mechanisms does not have to take into account biological plausibility. The only factors that would matter for this field are efficiency and reliability. It would also be useful to further expand readout mechanisms to not only detect contours, but also to identify shapes. Preliminary work could be done on shape identification by aiming to detect differences in the neural activity which arises when the model is presented with an S or an O-Shape contour. As results in this chapter have shown, contour detection performance differs between these two contour shapes. Thus, further characterizing the neural activity which gives rise to this performance difference might be a first step in devising a read-out mechanisms capable of performing shape identification.

In conclusion, the model presented in this chapter provides an advancement in the field of contour integration. To the knowledge of the author, there was no other model of contour integration which could integrate contours based on orientation cues **and** on spatial frequency cues, nor that it could account for their interactions. This model has presented putative mechanisms which may account for psychophysical results reported in the contour integration literature. Thus, it has help expand the knowledge on how the brain may process complex images and how it may perform the complex task of contour integration. Furthermore, it will provide aid in the development of novel psychophysics stimuli, and it might have applications in computer vision.

Chapter 5

Conclusion and Discussion

There are several levels at which visual perception can be studied. The physiological level provides an understanding of the biological mechanisms involved in visual perception. The behavioural level provides an understanding of the perceptual experience of subjects when presented with visual stimuli, and the theoretical level provides the means to connect empirical observations from these two fields. Only by combining these different levels of study can a full understanding of visual perception be achieved. In this thesis, by presenting interdisciplinary work in the field of contour integration, a significant contribution was made to the understanding of this basic visual process.

Although much is known about contour integration, important questions still remain to be answered before a full account of this topic can be achieved. Two of these open questions were addressed in this thesis.

The first question: "What are the effects of a dynamic history in a contour integration task, and how do top down processes affect contour integration in cases of high ambiguity". This question has been answered in Chapter 3 - Contour Integration in Dynamic Scenes.

Typically, in CI studies, stimuli are presented for a very brief period of time (>200ms), and the edge elements are static. Thus, in most of the previously employed stimuli for contour integration studies, motion, an integral part of natural vision, was missing. To address this issue, and to answer the question posed above, a set of experiments was conducted with a newly developed stimulus. Unlike stimuli used in most other CI tasks, these stimuli contained motion. Contours came into existence through the alignment of rotating gabor elements.

Prior to the introduction of these dynamic stimuli, it was impossible to evaluate contour integration in situations where motion was involved *and* oriented contours appeared due to the dynamics of edge elements. However, with this new paradigm, it was possible to determine whether the strong perceptual effects observed in CI studies carry over to visual stimuli containing motion. It has often been suggested that contour stimuli lead to "pop-out" percepts (Field et al., 1993).That is, to the experience of a strong perception of the outline formed by the aligned edge elements, which stands out from the randomly oriented surrounding elements. Pop-out percepts are believed to be stimulus driven (Connor et al., 2004), thus, to require little attentional resources

to be detected. However, the studies presented in Chapter 3 showed that in situations which require sustained attention for a prolong period of time, contour detection performance is impaired. As reported in Chapter 3 and in Grzymisch et al. (2017a), it was determined that subjects' ability to detect contours is impaired by whether or not a history period of at least 1.3s is presented prior to the appearance of a contour. In situations where no dynamic history is presented, subjects can detect contours with high accuracy (approximately 87% correct contour detection performance), however, when a history period is presented prior to the appearance of a contour, detection performance drops significantly (to approximately 67% correct). This finding implies that in dynamic scenes, top-down processes, such as attention, may play a key role in humans' ability to detect contours. This claim is supported by a growing body of evidence which suggests that the neural correlates of contour integration may not only be found in V1, but also in higher visual areas, which are known to be affected by attention (Lee and Nguyen, 2001; Mijovic et al., 2014). For a detailed discussion of this topic please refer to section 3.8.

In an experimental setting which made use of the newly developed stimuli and aimed to focus subjects' attentional resources, the question of how top down processes interact with CI under high ambiguity conditions was also explored. The results of these experiments indicate that in conditions in which subjects' ability to detect contours is impaired, top down influences are capable of partially restoring the ability to perceive contours. Once again, these findings are inline with electrophysiological evidence (Chen et al., 2014; Gilad et al., 2013), which shows that feedback loops are involved in CI. Thus, new evidence which challenges the commonly accepted view of contour integration as a pop-out process was presented, and a significant improvement in the methodology available to evaluate contour integration was introduced. To the knowledge of the author, this was the first paradigm in a CI task in which performance was starkly reduced by manipulating the history of a stimulus. In previous contour integration studies, performance has been shown to decrease when stimuli are presented for shorter periods of time (Beaudot and Mullen, 2001; Ernst et al., 2012); when contour element separation is increased (Mandon and Kreiter, 2005; May and Hess, 2008; Strother and Kubovy, 2006); when the deviations of contour elements' alignment to the contour's path is increased (Bex et al., 2001; Field et al., 1993); or if the phase difference of contour elements is increased (Hansen and Hess, 2006). Essentially, other contour integration paradigms have manipulated one, or more, physical characteristics of a stimulus in order to achieve different levels of contour visibility/detectability. This is undesirable if one seeks to study the neural correlates of observers' ability to detect contours. With the aforementioned methods, no distinction can be made on whether neural signatures differ due to the changes in the appearance of the stimulus, or due to the manipulation itself. However, this distinction can be made with the method presented in this thesis. Thus, this method will also prove valuable in the study of CI in this regard. For an in-depth discussion of the benefits which the method introduced will have in the contour integration field, please refer to section 3.8.

The second question: "Can a contour integration model which accounts for feature similarities (in terms of orientation and spatial frequency) be built to reproduce

psychophysical results. In particular, can the model be conceived around the principle that interactions of feature detectors are stronger if their preferred spatial frequencies are similar, rather than dissimilar, and if their RFs are visuotopically close, rather than far". This question has been answered in Chapter 4 - Feature Integration on Alignment and Spatial Frequency Similarities. Here, a structurally simple neuronal model of contour integration was devised. Based on the association field hypothesis, and on known mechanisms of the early visual cortex (V1), a mathematical framework was conceived to exploit the geometry and physical similarities of edge elements in stimuli used in CI experiments. The aim of this mathematical framework was to replicate the manner in which the visual system may make use of said properties. The model was based on a unifying theory capable of explaining a large pool of psychophysical data which characterized the interactions of orientation and spatial frequency cues in CI.

In the design of the model, several constrains based on the known biological properties of the visual system were put in place. The connection strengths between neural units were based on the association field proposed by Field et al. (1993), which has been shown to have biological basis (Gilbert and Wiesel, 1989; Shouval et al., 2000). The choice of function used to modulate connection strengths based on the differences in spatial frequency preference of neural detectors was also based on experimental evidence (Boucsein et al., 2011). The decision to scale the excitatory reach of lateral connection (of neural units) was supported by experimental evidence as well (Polat and Sagi, 1993; Tolhurst and Barfield, 1978), and by previous modelling studies (Ernst et al., 2016). Similarly, the read out mechanism employed was chosen not only because of the results which it yielded, but also because of its high likelihood to be biologically plausible (Li, 2014) (see section 4.7 for an in-depth discussion of these topics).

The above mentioned constraints were put in place because, by creating a model which processes stimuli in a similar manner to how the visual system processes them, it is possible to draw inferences on the mechanisms which give rise to visual perception. Thus, the model presented in Chapter 4 has furthered the understanding of these mechanisms, of how they may be tuned, and how they may interact with one-another.

To the knowledge of the author, Grzymisch et al. (2016) were the first to introduce a contour integration model capable of explaining the interactions of alignment and similarity cues. Through the replication of a large number of experimental results provided by Persike and Meinhardt (2015a,b); Persike et al. (2009), and through the replication of well established psychophysical phenomena (e.g.: effects of contour closure; detection performance with varying numbers of contour elements; contours defined by co-linear or parallel aligned elements - see section 4.6.2), the model has proven to be accurate and reliable. Thus, the introduction of this model is not only significant because of the advancements in understanding of the neural mechanisms which give rise to CI when alignment and similarity features interact, but also because of its predictive power. In section 4.6.3 the usefulness of the model as a predictive tool has been shown. As discussed in section 4.6.3 and 4.7 this model has the potential to help in the development of new stimuli to be used in psychophysical experiments, and it might also find applications in the field of computer vision.

Challenges: Although both the psychological experiments and the model reported

in this thesis present a success story, some challenges remain to be addressed. For the new psychophysical method introduced to study contour integration, the question of ecological validity remains to be answered. That is, do the stimuli developed share characteristics with scenes found in natural vision. Statistical analysis of natural images (Elder and Goldberg, 2002; Krüger, 1998; Sigman et al., 2001) has shown that the properties of the typical static stimuli used to study CI present similar statistical properties to those found in nature. Hence, if static, the stimuli employed would also be ecologically valid. However, the aim when developing this new stimulus was to introduce a form of motion to CI paradigms, and to ensure that the presence of a contour could *only* be deduced from alignment cues (since only then could CI be studied under dynamic conditions). Thus, statistical analysis of natural scenes containing motion still needs to be done in order to corroborate the ecological validity of these stimuli.

For the model, a detailed discussion of shortcomings and possible ways to address them is found in section 4.7. The most interesting of these is the issue of directionality. Two types of model variants were presented, one with a unidirectional projection preference, and one with a bidirectional projection preference. That is, in the model with a unidirectional projection preference, the lateral connections of neural populations only projected in one direction, whereas in the bidirectional model, projection preferences projected in polar opposite directions. The connection structure of the bidirectional model make it more biophysically plausible than those of the unidirectional model. Long range horizontal interactions expand isotropically in all directions (Bosking et al., 1997), thus the connection structure of the bidirectional model are more akin to those found in the visual system. While this is true, the bidirectional model did not yield results which matched (whether quantitatively or qualitatively) those found in psychophysics as closely as the results of the unidirectional model. As proposed by Schinkel-Bielefeld (2008), although the unidirectional connection structure may not be anatomically plausible, it could be supported by a refractory period. In this, a neuron would start to "view a contour in one direction", and after a certain refractory period the process of viewing the contour in the opposite direction would start. If this is the case, then the unidirectional connection structure would also be plausible. The root of this claim is that the dominating principle for grouping in the early visual development of children is not proximity, but rather coherent motion (Kellman and Shipley, 1991). Thus, the Gestalt rule of good continuation, which is believed to be the main Gestalt involved in CI (Field et al., 1993), could be learned from the grouping done based on coherent motion. Since motion is a directed process, it could be possible that contours are "viewed" in one direction, and then in another, making this process of viewing contours in polar opposite directions akin to the connection structure of the unidirectional model.

Outlook: In the work presented in Chapter 4 top down influences were not modelled. However, physiologically, it has been well established that in higher visual areas, receptive fields tend to be larger than in lower visual areas (Martinez and Alonso, 2003). These larger receptive fields are said to "see" a larger area of the visual field. Thus, while individual edge elements may excite single receptive fields in V1, it is likely

that receptive fields in V2, and higher areas, have a receptive field which can account for more than one edge element. If a sequence of edge elements is aligned, and thus generates a contour with an orientation which matches that of the receptive fields in higher visual areas, it is possible that these will have top-down influences in lower areas. In fact, Gilad et al. (2013) found that V1 cells show a response to individual gabor elements about 40-140ms after stimulus onset. However, only after 150-250ms did they see an increase in V1 responses due to the alignment of individual edge elements. This suggests that higher order visual areas exerted top-down influences, which increase the response of cells in V1. As such, it might be interesting to model this kind of influence in the future.

Modelling the effects of top-down influences would not only make the model presented in this thesis more complete, but it would also allow for the modelling of other interesting conditions. Mainly, the model could be extended to account for stimuli similar to those employed in the psychophysical experiments reported in Chapter 3 (Contour Integration in Dynamic Scenes). If neural modulation of activity is added on the basis of transient attention and sustained attention, then the model could also process dynamic stimuli. It should also be noted that with dynamic stimuli it would be necessary to account for the changing stimulus. Thus, neural populations would have to be adapted to account for different orientations, as now they are only tuned to the specific orientation which stimulates them. This could be done by modelling orientation columns rather than populations which are tuned to one specific orientation.

The effects of both transient and sustained attention have been studied in the past (Carrasco et al., 2006; Yeshurun and Carrasco, 1998; Yeshurun and Levy, 2003; Yeshurun et al., 2008). Transient attention, as the name suggests, is short-lived and is typically caused by a change in a stimulus that attracts attention. Sustained attention is generally and broadly defined as the ability to direct and focus cognitive activity on a specific task/stimulus (DeGangi and Porges, 1990). Unlike transient attention, sustained attention is long-lived, and it is typically voluntary. Since transient attention is short-lived, and often caused by changes in stimuli which attract attention, it could be modelled with a function which decays over time (with a rapid rate of decay, which would make the effects of this modulation relevant only for a short period of time). Similarly, sustained attention could also be modelled with a function which decays over time. However, with a slow rate of decay. The slower rate of decay would model the loss in ability to sustain attention for prolonged periods of time. These two modulating forces could target excitatory connections via a multiplicative relationship. If this is the case, and if the functions are calibrated appropriately, excitatory connections could be amplified by transient attention up to a time point $T^{Balance}_{Transient}$, and beyond that point either decreased, or left to evolve (virtually) without any effects from this modulation (if the decay function approaches 1 asymptotically beyond $T^{Balance}_{Transient}$). Similarly, sustained attention could provide the same type of modulation, up to a different time point $T^{Balance}_{Sustained}$.

The beneficial effects of the cues employed in the experiment reported in Chapter 3 (Contour Integration in Dynamic Scenes) could also be modelled with the aid of decaying functions as models of both transient and sustained attention. An offset could be placed on the functions to simulate cued trials, such that in cued trials the offset

would increase the maximum of the decay function. Thus, not only would the effects of both of these attentional modulation modalities last longer, but they would also be stronger in cued trials. Finally, in order to model perceptual learning (a common phenomenon in psychophysics and also observed in the results reported in Chapter 3), a function which increases the strength of excitatory connections based on the amount of trials to which the system has been exposed could also be implemented.

This thesis has presented work which advanced the state of knowledge in contour integration. Important questions in two distinct areas of study in this field have been answered. Nevertheless, both, the new paradigm designed introduced in Chapter 3, and the proposed model present many future opportunities. Of most interest for the new paradigm would be to perform neurophysiological recordings of V1 (and higher cortical areas), while an animal is presented with the newly developed dynamic stimuli, in which contours appear a few seconds after stimulus onset. This would allow for the effects of the presentation of a contour stimulus to be studied without the sudden onset of a transient in neural activity, thus, to isolate the effects of viewing contours in the visual cortex. For the model, first, the suggestions proposed in the conclusion of Chapter 4 (Feature Integration on Alignment and Spatial Frequency Similarities) should be implemented. Second, the model could be extended to account for top-down influences in the manner described above. This would allow to explore contour integration in a more complete manner, and be at the forefront of new contour integration studies which are bound to include dynamic components in the future.

Supplementary Material

For a sample of the stimuli used in the experiments please refer to the supplementary materials section of Grzymisch et al. (2017a). The supplementary materials can be accessed directly via this link, or via the URL: `http://journal.frontiersin.org/article/10.3389/fpsyg.2017.01501/full?&utm_source=Email_to_authors_&utm_medium=Email&utm_content=T1_11.5e1_author&utm_campaign=Email_publication&field=&journalName=Frontiers_in_Psychology&id=267579`.

- Video 1: Peak condition, with added markers indicating contour location
- Video 2: Peak condition, no markers
- Video 3: Long condition, with added markers indicating contour location
- Video 4: Long condition, no markers

Bibliography

Adams, D. L. and Horton, J. C. (2003). A precise retinotopic map of primate striate cortex generated from the representation of angioscotomas. *Journal of Neuroscience*, 23(9):3771–3789.

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

Angelucci, A., Levitt, J. B., Walton, E. J. S., Hupé, J.-M., Bullier, J., and Lund, J. S. (2002). Circuits for local and global signal integration in primary visual cortex. *Journal of Neuroscience*, 22(19):8633–8646.

Anstis, S., Verstraten, F. A. J., and Mather, G. (1998). The motion aftereffect. *Trends in Cognitive Sciences*, 2:111–117.

Awh, E., Armstrong, K. M., and Moore, T. (2006). Visual and oculomotor selection: links, causes and implications for spatial attention. *Trends in Cognitive Sciences*, 10:124–130.

Bacon-Mace, N., Mace, M. J. M., Fabre-Thorpe, M., and Thorpe, S. J. (2005). The time course of visual processing: Backward masking and natural scene categorisation. *Vision Research*, 45(11):1459–1469.

Basso, M. A. and Wurtz, R. H. (1997). Modulation of neuronal activity by target uncertainty. *Nature*, 389(6646):66.

Bauer, R. and Heinze, S. (2002). Contour integration in striate cortex. *Experimental Brain Research*, 147(2):145–152.

Beaudot, W. H. and Mullen, K. T. (2001). Processing time of contour integration: the role of colour, contrast, and curvature. *Perception*, 30(7):833–853.

Beck, J., Rosenfeld, A., and Ivry, R. (1989). Line segregation. *Spatial Vision*, 4(2):75–101.

Bex, P. J., Simmers, A. J., and Dakin, S. C. (2001). Snakes and ladders: The role of temporal modulation in visual contour integration. *Vision Research*, 42:653–659.

Bosking, W. H., Zhang, Y., Schofield, B., and Fitzpatrick, D. (1997). Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex. *Journal of Neuroscience*, 17(6):2112–2127.

Boucsein, C., Nawrot, M., Schnepel, P., and Aertsen, A. (2011). Beyond the cortical column: abundance and physiology of horizontal connections imply a strong role for inputs from the surround. *Frontiers in Neuroscience*, 5:32.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10:433–436.

Braun, J. (1999). On the detection of salient contours. *Spatial Vision*, 12(2):211–225.

Brown, S. and Schafer, E. A. (1888). An investigation into the functions of the occipital and temporal lobes of the monkey's brain. *Philosophical Transactions of the Royal Society of London*, 179:303–327.

Burr, D. C. and Santoro, L. (2001). Temporal integration of optic flow, measured by contrast and coherence thresholds. *Vision Research*, 41(15):1891–1899.

Carandini, M. (2000). Visual cortex: Fatigue and adaptation. *Current Biology*, 10(16):605–607.

Carandini, M. (2012). Area V1. *Scholarpedia*, 7(7):12105.

Carandini, M., Heeger, D. J., and Movshon, J. A. (1997). Linearity and Normalization in Simple Cells of the Macaque Primary Visual Cortex. *The Journal of Neuroscience*, 17(21):8621–8644.

Carrasco, M., Loula, F., and Ho, Y. X. (2006). How attention enhances spatial resolution: Evidence from selective adaptation to spatial frequency. *Attention, Perception, & Psychophysics*, 68(6):1004–1012.

Castellano, M., Ploechl, M., Vicente, R., and Pipa, G. (2014). Neuronal oscillations during contour integration of dynamic visual stimuli form parietal/frontal networks. *Frontiers in Integrative Neuroscience*, 8(64).

Chelazzi, L., Miller, E. K., Duncan, J., and Desimone, R. (1993). A neural basis for visual search in inferior temporal cortex. *Nature*, 363:345–347.

Chen, M., Yan, Y., Gong, X., Gilbert, C. D., Liang, H., and Li, W. (2014). Incremental Integration of Global Contours through Interplay between Visual Cortical Areas. *Neuron*, 82(3):682–694.

Clarke, S. and Miklossy, J. (1990). Occipital cortex in man: Organization of callosal connections, related myelo- and cytoarchitecture, and putative boundaries of functional visual areas. *The Journal of Comparative Neurology*, 298(2):188–214.

Cleland, B. G. and Levick, W. (1974). Brisk and sluggish concentrically organized ganglion cells in the cat's retina. *The Journal of Physiology*, 240:421–456.

Connor, C. E., Egeth, H. E., and Yantis, S. (2004). Visual attention: bottom-up versus top-down. *Current Biology*, 14(19):R850–R852.

Cope, D., Blakeslee, B., and McCourt, M. E. (2009). Simple cell response properties imply receptive field structure: Balanced gabor and/or bandlimited field functions. *Journal of the Optical Society of America. Optics, image science, and vision*, 26:2067–2092.

Coren, S. and Girgus, J. S. (1980). Principles of perceptual organization and spatial distortion: The gestalt illusions. *Journal of Experimental Psychology: Human Perception and Performance*, 6(3):404–412.

Cowan, J. D., Neuman, J., and van Drongelen, W. (2016). Wilson-Cowan equations for neocortical dynamics. *The Journal of Mathematical Neuroscience*, 6(1):1.

Curcio, C. A. and Allen, K. A. (1990). Topography of ganglion cells in human retina. *Journal of Comparative Neurology*, 300(1):5–25.

Dakin, S. C. and Hess, R. F. (1998). Spatial-frequency tuning of visual contour integration. *Journal of the Optical Society of America A.*, 15(6):1486–1499.

De Valois, R. L., Albrecht, D. G., and Thorell, L. G. (1982). Spatial frequency selectivity of cells in macaque visual cortex. *Vision research*, 22(5):545–559.

DeGangi, G. and Porges, S. (1990). Neuroscience foundations of human performance. *Rockville, MD: American Occupational Therapy Association Inc.*

Desimone, R. and Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, 18:193–222.

Dowling, J. (2009). Retina: An overview. In Squire, L. R., editor, *Encyclopedia of Neuroscience*, pages 159–169. Academic Press, Oxford.

Drewes, J., Zhu, W., Wutz, A., and Melcher, D. (2015). Dense sampling reveals behavioral oscillations in rapid visual categorization. *Scientific reports*, 5.

Elder, J. H. and Goldberg, R. M. (2002). Ecological statistics of gestalt laws for the perceptual organization of contours. *Journal of Vision*, 2(4):5–5.

Enroth-Cugell, C. and Robson, J. G. (1966). The contrast sensitivity of retinal ganglion cells of the cat. *The Journal of Physiology*, 187(3):517–552.

Ernst, M. O. and Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415:429–433.

Ernst, U. A., Mandon, S., Schinkel-Bielefeld, N., Neitzel, S. D., Kreiter, A. K., and Pawelzik, K. R. (2012). Optimality of human contour integration. *PLOS Computational Biology*, 8(5):1–17.

Ernst, U. A., Schiffer, A., Persike, M., and Meinhardt, G. (2016). Contextual interactions in grating plaid configurations are explained by natural image statistics and neural modeling. *Frontiers in Systems Neuroscience*, 10.

Faisal, A. A., Selen, L. P., and Wolpert, D. M. (2008). Noise in the nervous system. *Nature reviews. Neuroscience*, 9(4):292.

Fatt, P. and Katz, B. (1952). Spontaneous subthreshold activity at motor nerve endings. *The Journal of Physiology*, 117(1):109–128.

Feldman, J. and Singh, M. (2005). Information along contours and object boundaries. *Psychological Review*, 112(1):243.

Field, D. J., Hayes, A., and Hess, R. F. (1993). Contour integration by the human visual system: Evidence for a local association field. *Vision Research*, 33(2):173–193.

Fischer, B. and Breitmeyer, B. (1987). Mechanisms of visual attention revealed by saccadic eye movements. *Neuropsychologia*, 25(1):73–83.

Fisher, D. L. (1984). Central capacity limits in consistent mapping, visual search tasks: Four channels or more? *Cognitive Psychology*, 16(4):449–484.

Galashan, F. O., Sassen, H. C., Kreiter, A. K., and Wegener, D. (2013). Monkey Area MT Latencies to Speed Changes Depend on Attention and Correlate with Behavioral Reaction Times. *Neuron*, 78(4):740–750.

Geisler, W. S., Perry, J. S., Super, B., and Gallogly, D. (2001). Edge co-occurrence in natural images predicts contour grouping performance. *Vision research*, 41(6):711–724.

Gilad, A., Meirovithz, E., and Slovin, H. (2013). Population Responses to Contour Integration: Early Encoding of Discrete Elements and Late Perceptual Grouping. *Neuron*, 78(2):389–402.

Gilbert, C. D. and Li, W. (2013). Top-down influences on visual processing. *Nature Reviews Neuroscience*, 14:350–363.

Gilbert, C. D., Li, W., and Piech, V. (2009). Perceptual learning and adult cortical plasticity. *The Journal of Physiology*, 587:2743–2751.

Gilbert, C. D. and Wiesel, T. N. (1989). Columnar specificity of intrinsic horizontal and corticocortical connections in cat visual cortex. *Journal of Neuroscience*, 9(7):2432–2442.

Goldberg, M. E. and Wurtz, R. H. (1972). Activity of superior colliculus in behaving monkey. II. Effect of attention on neuronal responses. *Journal of Neurophysiology*, 35(4):560–574.

Goldstein, E. B. (2009). *Sensation and Perception*. Cengage Learning.

Grzymisch, A., Grimsen, C., and Ernst, U. A. (2013). Contour integration in static and dynamic scenes. *Perception 42 ECVP Abstract Supplement*.

Grzymisch, A., Grimsen, C., and Ernst, U. A. (2015). Attentional effects in contour integration in dynamic scenes. *Vision Science Society (VSS) Annual Meeting*, page Poster 26.4077.

Grzymisch, A., Grimsen, C., and Ernst, U. A. (2017a). Contour integration in dynamic scenes: Impaired detection performance in extended presentations. *Frontiers in Psychology*, 8:1501.

Grzymisch, A., Persike, M., and Ernst, U. A. (2016). Modelling the effects of spatial frequency jitters in a contour integration paradigm. *Perception 45 ECVP Abstract Supplement*, page Poster 1P072.

Grzymisch, A., Persike, M., and Ernst, U. A. (2017b). Contour integration in multiple features explained in a recurrent network model. *Perception 46 ECVP Abstract Supplement*.

Grzymisch, A., Schiffer, A., Persike, M., and Ernst, U. (2017c). Integration of orientation and spatial frequency in a model of visual cortex. *Bernstein Conference 2017*.

Haenny, P. E., , and Schiller, P. H. (1988). State dependent activity in monkey visual cortex. *Experimental Brain Research*, 69(2):225–244.

Hamilton, D. B., Albrecht, D. G., and Geisler, W. S. (1989). Visual cortical receptive fields in monkey and cat: spatial and temporal phase transfer function. *Vision Research*, 29(10):1285–1308.

Hansen, B. C. and Hess, R. F. (2006). The role of spatial phase in texture segmentation and contour integration. *Journal of Vision*, 6(5):594–615.

Hansen, T. and Neumann, H. (2008). A recurrent model of contour integration in primary visual cortex. *Journal of Vision*, 8(8):8.

Hartline, H. K. (1938). The response of single optic nerve fibers of the vertebrate eye to illumination of the retina. *American Journal of Physiology–Legacy Content*, 121(2):400–415.

Hebb, D. O. (1949). *The organization of behavior: A neuropsychological approach*. John Wiley & Sons.

Heider, F. (1970). Gestalt theory: Early history and reminiscences. *Journal of the History of the Behavioral Sciences*, 6:131–139.

Herzog, M. H., Ernst, U. A., Etzold, A., and Eurich, C. W. (2003). Local interactions in neural networks explain global effects in gestalt processing and masking. *Neural Computation*, 15(9):2091–2113.

Herzog, M. H., Koch, C., and Fahle, M. (2001). Shine-through: temporal aspects. *Vision Research*, 41(18):2337–2346.

Hess, R. and Field, D. J. (1999). Integration of contours: new insights. *Trends in Cognitive Sciences*, 3(12):480–486.

Hess, R. F., Beaudot, W. H. A., and Mullen, K. T. (2001). Dynamics of contour integration. *Vision Research*, 41:1023–1037.

Hess, R. F., Hayes, A., and Field, D. J. (2003). Contour integration and cortical processing. *Journal of Physiology-Paris*, 97(2–3):105–119.

Hubel, D. (1995). *Eye, Brain, and Vision*. Scientific American Library Series. Henry Holt and Company.

Hubel, D. H. and Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *The Journal of Physiology*, 148:574–591.

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*, 160:106–154.

Hubel, D. H. and Wiesel, T. N. (1974). Sequence regularity and geometry of orientation columns in the monkey striate cortex. *The Journal of Comparative Neurology*, 158(3):267–293.

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

Institute, Q. B. (2016). The four lobes of the cerebral cortex.

Jonides, J. and Yantis, S. (1988). Uniqueness of abrupt visual onset in capturing attention. *Perception & Psychophysics*, 43(4):346–354.

Joseph, J. S., Chun, M. M., and Nakayama, K. (1997). Attentional requirements in a preattentive feature search task. *Nature*, 387:805–807.

Kandel, E. R., Schwartz, J. H., and Jessell, T. M. (1991). *Principles of Neural Science*. Appleton and Lange, 25 Van Zant Streer, East Norwalk, Connecticut 06855, USA, 3 edition.

Kellman, P. J. and Shipley, T. F. (1991). A theory of visual interpolation in object perception. *Cognitive Psychology*, 23(2):141–221.

Kleiner, M., Brainard, D., and Pelli, D. M. (2007). What's new in Psychtoolbox-3? *Perception 36 ECVP Abstract Supplement*.

Kleinschmidt, A., Buechel, C., Hutton, C., Friston, K. J., and Frackowiak, R. S. (2002). The neural structures expressing perceptual hysteresis in visual letter recognition. *Neuron*, 34(4):659–666.

Koffka, K. (1935). *Principles of Gestalt Psychology*. Lund Humphries, London.

Kohn, A. (2007). Visual Adaptation: Physiology, Mechanisms, and Functional Benefits. *Journal of Neurophysiology*, 97(5):3155–3164.

Kovacs, I. (1996). Gestalten of today: early processing of visual contours and surfaces. *Behavioural Brain Research*, 82(1):1–11.

Kovacs, I. and Julesz, B. (1993). A closed curve is much more than an incomplete one: Effect of closure in figure-ground segmentation. *Proceedings of the National Academy of Sciences*, 90(16):7495–7497.

Krüger, N. (1998). Collinearity and parallelism are statistically significant second-order relations of complex cell responses. *Neural Processing Letters*, 8(2):117–129.

Kubovy, M., Holcombe, A. O., and Wagemans, J. (1998). On the lawfulness of grouping by proximity. *Cognitive Psychology*, 35(1):71–98.

Kuffler, S. W. (1953). Discharge patterns and functional organization of mammalian retina. *Journal of Neurophysiology*, 16(1):37–68.

Lamy, D. and Egeth, H. E. (2003). Attentional capture in singleton-detection and feature-search modes. *Journal of Experimental Psychology: Human Perception and Performance*, 29(5):1003.

Ledgeway, T., Hess, R. F., and Geisler, W. S. (2005). Grouping local orientation and direction signals to extract spatial contours: Empirical tests of "association field" models of contour integration. *Vision Research*, 45(19):2511–2522.

Lee, B. B., Pokorny, J., Martin, P. R., Valbergt, A., and Smith, V. C. (1990). Luminance and chromatic modulation sensitivity of macaque ganglion cells and human observers. *Journal of the Optical Society of America A*, 7(12):2223–2236.

Lee, T. S. and Nguyen, M. (2001). Dynamics of subjective contour formation in the early visual cortex. *PNAS*, 98:1907–1911.

Li, W. and Gilbert, C. D. (2002). Global Contour Saliency and Local Colinear Interactions. *Journal of Neurophysiology*, 88(5):2846–2856.

Li, W., Piech, V., and Gilbert, C. D. (2008). Learning to Link Visual Contours. *Neuron*, 57(3):442–451.

Li, Y. and Li, S. (2015). Contour integration, attentional cuing, and conscious awareness: An investigation on the processing of collinear and orthogonal contours. *Journal of vision*, 15(16):10–10.

Li, Z. (1998). A neural model of contour integration in the primary visual cortex. *Neural Computation*, 10:903–940.

Li, Z. (1999). Contextual influences in V1 as a basis for pop out and asymmetry in visual search. *Proceedings of the National Academy of Sciences of the United States of America*, 98:10530–10535.

Li, Z. (2014). *Understanding Vision: Theory, Models, and Data*. Oxford University Press.

Luck, S. J., Chelazzi, L., Hillyard, S. A., and Desimone, R. (1997). Neural Mechanisms of Spatial Selective Attention in Areas V1, V2, and V4 of Macaque Visual Cortex. *Journal of Neurophysiology*, 77(1):24–42.

Mandon, S. and Kreiter, A. K. (2005). Rapid contour integration in macaque monkeys. *Vision Research*, 45:291–300.

Martinez, L. M. and Alonso, J.-M. (2003). Complex receptive fields in primary visual cortex. *The Neuroscientist*, 9(5):317–331.

Mathes, B. and Fahle, M. (2007). The electrophysiological correlate of contour integration is similar for color and luminance mechanisms. *Psychophysiology*, 44(2):305–322.

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

May, K. and Hess, R. (2008). Effects of element separation and carrier wavelength on detection of snakes and ladders: Implications for models of contour integration. *Journal of Vision*, 8:1–23.

McAdams, C. J. and Maunsell, J. H. (1999a). Effects of Attention on the Reliability of Individual Neurons in Monkey Visual Cortex. *Neuron*, 23(4):765–773.

McAdams, C. J. and Maunsell, J. H. R. (1999b). Effects of Attention on Orientation-Tuning Functions of Single Neurons in Macaque Cortical Area V4. *Journal of Neuroscience*, 19(1):431–441.

McIlhagga, W. H. and Mullen, K. T. (1996). Contour integration with colour and luminance contrast. *Vision Research*, 36(9):1265–1279.

McLaughlin, D., Shapley, R., Shelley, M., and Wielaard, D. J. (2005). A neuronal network model of macaque primary visual cortex (V1): Orientation selectivity and dynamics in the input layer 4C. *PNAS*, 97(14):8087–8092.

Melcher, D., Crespi, S., Bruno, A., and Morrone, M. C. (2004). The role of attention in central and peripheral motion integration. *Vision research*, 44(12):1367–1374.

Mijovic, B., Vos, M. D., Vanderperren, K., Machilsent, B., Sunaert, S., Huffel, S. V., and Wagemans, J. (2014). The dynamics of contour integration: A simultaneous EEG-fMRI study. *Neuroimage*, 88:10–21.

Moran, J. and Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science*, 229(4715):782–784.

Movshon, A. J. and Simoncelli, E. P. (2015). Representation of naturalistic image structure in the primate visual cortex. *Cold Spring Harbor Symposia on Quantitative Biology*.

Müller, J. R., Philiastides, M. G., and Newsome, W. T. (2005). Microstimulation of the superior colliculus focuses attention without moving the eyes. *Proceedings of the National Academy of Sciences of the United States of America*, 102(3):524–529.

Nienborg, H., Bridge, H., Parker, A. J., and Cumming, B. G. (2004). Receptive field size in V1 neurons limits acuity for perceiving disparity modulation. *Journal of Neuroscience*, 24(9):2065–2076.

Nugent, A. K., Keswani, R. N., Woods, R. L., and Peli, E. (2003). Contour integration in peripheral vision reduces gradually with eccentricity. *Vision Research*, 43(23):2427–2437.

Oyama, T. (1961). Perceptual grouping as a function of proximity. *Perceptual and Motor Skills*, 13(3):305–306.

Palmer, S. E. (1999). *Vision Science Photons to Phenomenology*. Bradford Book.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*, 10:437–442.

Persike, M. and Meinhardt, G. (2015a). Cue combination anisotropies in contour integration: The role of lower spatial frequencies. *Journal of Vision*, 15(5):17.

Persike, M. and Meinhardt, G. (2015b). Effects of spatial frequency similarity and dissimilarity on contour integration. *PLOS ONE*, 10(6):1–19.

Persike, M. and Meinhardt, G. (2016). Contour integration with corners. *Vision research*, 127:132–140.

Persike, M., Olzak, L. A., and Meinhardt, G. (2009). Contour integration across spatial frequency. *Journal of Experimental Psychology: Human Perception and Performance*, 35:1629–1648.

Pettet, M. W., McKee, S. P., and Grzywacz, N. M. (1998). Constraints on long range interactions mediating contour detection. *Vision research*, 38(6):865–879.

Polat, U., Mizobe, K., Pettet, M. W., and Norcia, T. K. A. M. (1998). Collinear stimuli regulate visual responses depending on cell's contrast threshold. *Nature*, 391:580–584.

Polat, U. and Sagi, D. (1993). Lateral interactions between spatial channels: suppression and facilitation revealed by lateral masking experiments. *Vision research*, 33(7):993–999.

Purves, D. (2004). *Neuroscience*. The Japanese classic collection. Sinauer Associates.

Reynolds, J. H., Chelazzi, L., and Desimone, R. (1999). Competitive Mechanisms Subserve Attention in Macaque Areas V2 and V4. *Journal of Neuroscience*, 19(5):1736–1753.

Rock, I. and Brosgole, L. (1964). Grouping based on phenomenal proximity. *Journal of Experimental Psychology*, 67(6):531.

Sakamoto, K., Nakajima, H., Suzuki, T., and Yano, M. (2008). A "global closure" effect in contour integration. In *International Conference on Neural Information Processing*, pages 259–266. Springer.

Schinkel-Bielefeld, N. (2008). *Contour Integration Models Predicting Human Behavior*. PhD thesis.

Schoups, A. A., Vogels, R., Qian, N., and Orban, G. (2001). Practising orientation identification improves orientation coding in V1 neurons. *Nature*, 412:549–553.

Schwiedrzik, C. M., Leitner, F., Singer, W., Ruff, C. C., and Melloni, L. (2011). Perceptual hysteresis and adaptation are expressed in distinct cortical networks. *Frontiers in Human Neuroscience. Conference Abstract: XI International Conference on Cognitive Neuroscience (ICON XI)*, 207(51).

Schwiedrzik, C. M., Ruff, C. C., Lazar, A., Leitner, F. C., Singer, W., and Melloni, L. (2012). Untangling perceptual memory: Hysteresis and adaptation map into separate cortical networks. *Cerebral Cortex*, 24(5):1152.

Shiu, L.-P., , and Pashler, H. (1992). Improvement in line orientation discrimination is retinally local but dependent on cognitive set. *Perception & Psychophysics*, 52(5):582–588.

Shmuel, A., Korman, M., Sterkin, A., Harel, M., Ullman, S., Malach, R., and Grinvald, A. (2005). Retinotopic axis specificity and selective clustering of feedback projections from V2 to V1 in the owl monkey. *Journal of Neuroscience*, 25(8):2117–2131.

Shouval, H. Z., Goldberg, D. H., Jones, J. P., Beckerman, M., and Cooper, L. N. (2000). Structured long-range connections can provide a scaffold for orientation maps. *Journal of Neuroscience*, 20(3):1119–1128.

Sigman, M., Cecchi, G. A., Gilbert, C. D., and Magnasco, M. O. (2001). On a common circle: natural scenes and gestalt rules. *Proceedings of the National Academy of Sciences*, 98(4):1935–1940.

Smits, J. T. and Vos, P. G. (1987). The perception of continuous curves in dot stimuli. *Perception*, 16(1):121–131.

Solomon, J. A., Watson, A. B., and Morgan, M. J. (1999). Transducer model produces facilitation from opposite-sign flanks. *Vision Research*, 39(5):987–992.

Spitzer, H. and Hochstein, S. (1985). A complex-cell receptive-field model. *Journal of Neurophysiology*, 53(5):1266–1286.

Stahl, J. S. and Wang, S. (2008). Globally optimal grouping for symmetric closed boundaries by combining boundary and region information. *IEEE Transactions on pattern analysis and machine intelligence*, 30(3):395–411.

Stern, E. A., Kincaid, A. E., and Wilson, C. J. (1997). Spontaneous subthreshold membrane potential fluctuations and action potential variability of rat corticostriatal and striatal neurons in vivo. *Journal of Neurophysiology*, 77(4):1697–1715.

Stettler, D. D., Das, A., Bennett, J., and Gilbert, C. D. (2002). Lateral connectivity and contextual interactions in macaque primary visual cortex. *Neuron*, 36(4):739–750.

Strother, L. and Kubovy, M. (2006). On the surprising salience of curvature in grouping by proximity. *Journal of Experimental Psychology: Human Perception and Performance*, 32(5):226–234.

Talbot, S. and Marshall, W. (1941). Physiological studies on neural mechanisms of visual localization and discrimination. *American Journal of Ophthalmology*, 24(11):1255–1264.

Tehovnik, E. J., Slocum, W. M., and Schiller, P. H. (2003). Saccadic eye movements evoked by microstimulation of striate cortex. *European Journal of Neuroscience*, 17(4):870–878.

Theeuwes, J. (1991). Exogenous and endogenous control of attention: The effect of visual onsets and offsets. *Attention, Perception, & Psychophysics*, 49(1):83–90.

Theeuwes, J., De Vries, G.-J., and Godijn, R. (2003). Attentional and oculomotor capture with static singletons. *Attention, Perception, & Psychophysics*, 65(5):735–746.

Todorovic, D. (2011). What is the origin of the gestalt principles? *Humana Mente*, pages 1–19.

Tolhurst, D. and Barfield, L. (1978). Interactions between spatial frequency channels. *Vision Research*, 18(8):951–958.

Treue, S. and Trujillo, J. C. M. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature*, 399:575–579.

Tsodyks, M., Pawelzik, K., and Markram, H. (1998). Neural Networks with Dynamic Synapses. *Neural Computation*, 10:821–835.

Tversky, T., Geisler, W. S., and Perry, J. S. (2004). Contour grouping: Closure effects are explained by good continuation and proximity. *Vision Research*, 44(24):2769–2777.

Ursino, M. and La Cara, G. E. (2004). A model of contextual interactions and contour detection in primary visual cortex. *Neural Networks*, 17(5):719–735.

Vancleef, K. and Wagemans, J. (2013). Component processes in contour integration: A direct comparison between snakes and ladders in a detection and a shape discrimination task. *Vision Research*, 92:39–46.

VanRullen, R., Delorme, A., and Thorpe, S. (2001). Feed-forward contour integration in primary visual cortex based on asynchronous spike propagation. *Neurocomputing*, 38–40:1003–1009.

Visscher, K. M., Miezin, F. M., Kelly, J. E., Buckner, R. L., Donaldson, D. I., McAvoy, M. P., Bhalodia, V. M., and Petersen, S. E. (2003). Mixed blocked/event-related designs separate transient and sustained activity in fMRI. *NeuroImage*, 19(4):1694–1708.

Wagemans, J. (2015). *The Oxford Handbook of Perceptual Organization*. Oxford library of psychology. Oxford University Press.

Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., and von der Heydt, R. (2012a). A century of gestalt psychology in visual perception: I. Perceptual grouping and figure–ground organization. *Psychological Bulletin*, 138(6):1172.

Wagemans, J., Feldman, J., Gepshtein, S., Kimchi, R., Pomerantz, J. R., van der Helm, P. A., and van Leeuwen, C. (2012b). A century of gestalt psychology in visual perception: II. Conceptual and theoretical foundations. *Psychological Bulletin*, 138(6):1218.

Wandell, B. A. (1995). *Foundations of Vision*. Sinauer Associates.

Wertheimer, M. (1923). Untersuchungen zur Lehre der Gestalt II. *Psychologische Forschung 4*, pages 301–350. [Translation published as: Laws of organization in perceptual forms, in A Source Book of Gestalt Psychology Ed. W Ellis (1938, London: Routledge and Kegan Paul) pp 71-88].

Wiesel, T. N. (1960). Receptive fields of ganglion cells in the cat's retina. *The Journal of Physiology*, 153:583–594.

Williams, C. B. and Hess, R. F. (1998). Relationship between facilitation at threshold and suprathreshold contour integration. *JOSA A*, 15(8):2046–2051.

Williams, L. R. and Thornber, K. K. (2001). Orientation, scale, and discontinuity as emergent properties of illusory contour shape. *Neural Computation*, 13(8):1683–1711.

Wilson, H. R. and Cowan, J. D. (1972). Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical Journal*, 12:1–24.

Woods, R. L., Nugent, A. K., and Peli, E. (2002). Lateral interactions: Size does matter. *Vision research*, 42(6):733–745.

Wutz, A. and Melcher, D. (2013). Temporal buffering and visual capacity: The time course of object formation underlies capacity limits in visual cognition. *Attention, Perception, & Psychophysics*, 75(5):921–933.

Yen, S.-C. and Finkel, L. H. (1998). Extraction of perceptually salient contours by striate cortical networks. *Vision research*, 38(5):719–741.

Yeshurun, Y. and Carrasco, M. (1998). Attention improves or impairs visual performance by enhancing spatial resolution. *Nature*, 396(6706).

Yeshurun, Y. and Levy, L. (2003). Transient spatial attention degrades temporal resolution. *Psychological Science*, 14(3):225–231.

Yeshurun, Y. and Marom, G. (2008). Transient spatial attention and the perceived duration of brief visual events. *Visual Cognition*, 16(6):826–848.

Yeshurun, Y., Montagna, B., and Carrasco, M. (2008). On the flexibility of sustained attention and its effects on a texture segmentation task. *Vision research*, 48(1):80–95.

Appendices

Appendix A

A.1 Figures

Figures found in Persike and Meinhardt (2015a,b); Persike et al. (2009) which might aid the reader's understanding of these experiments are found in this section.

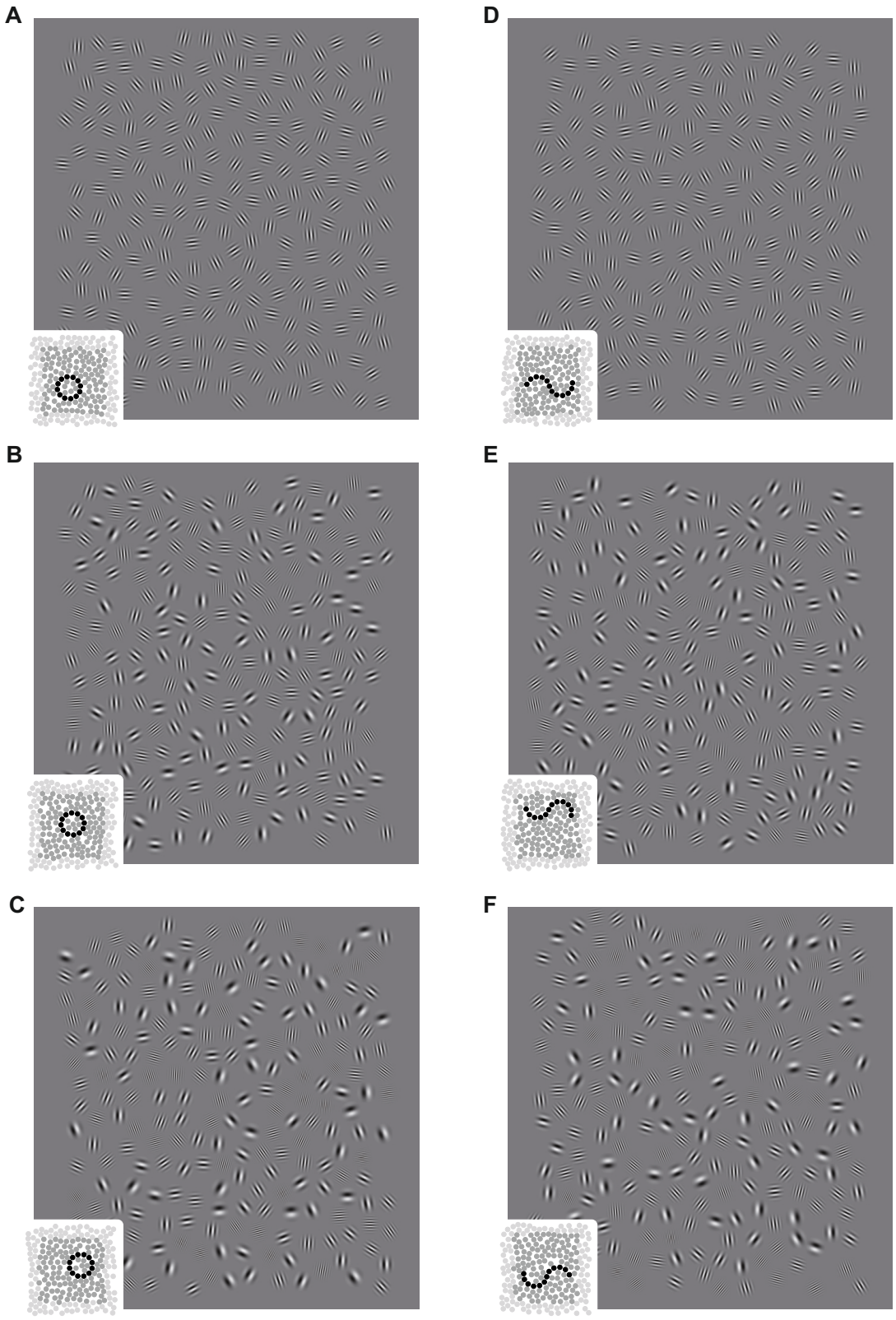

Figure A.1: Example of target display stimuli for (Persike et al., 2009) Experiment One with a 0° tilt angle jitter in the contour path. The figures contained in the boxes on the bottom left corner of each panel indicate the position and shape of the contour in the larger stimulus display. Left panels O-Shape contours, right panels S-Shape contours. Spatial frequency jitter levels: UNI (**A** & **D**); MEDIUM (**B** & **E**); HIGH (**C** & **F**). Image obtained from (Persike et al., 2009).

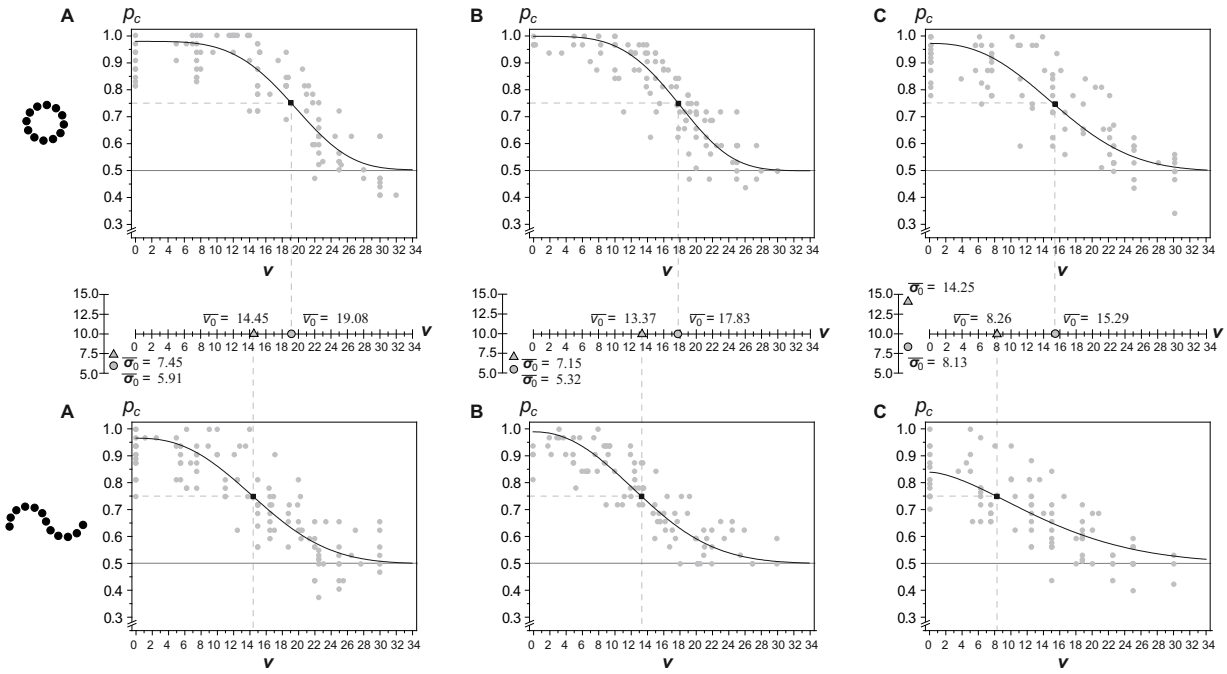

Figure A.2: Results from (Persike et al., 2009) Experiment One. Panels **A**, **B**, and **C** O-Shape stimulus results. Panels **D**, **E**, and **F** S-Shape stimulus results. Left column UNI spatial frequency jitter; middle column MEDIUM spatial frequency jitter; right column HIGH spatial frequency jitter. Fitted psychometric curves with a Weibull function that intersects at the between subject mean tilt angle threshold of $V = 0.75$ and has a slop calculated from the mean standard deviation estimate are shown in solid black lines in each of the panels. Mean jitter threshold standard deviations are shown in an axis in between the two rows of panels. Image obtained from (Persike et al., 2009).

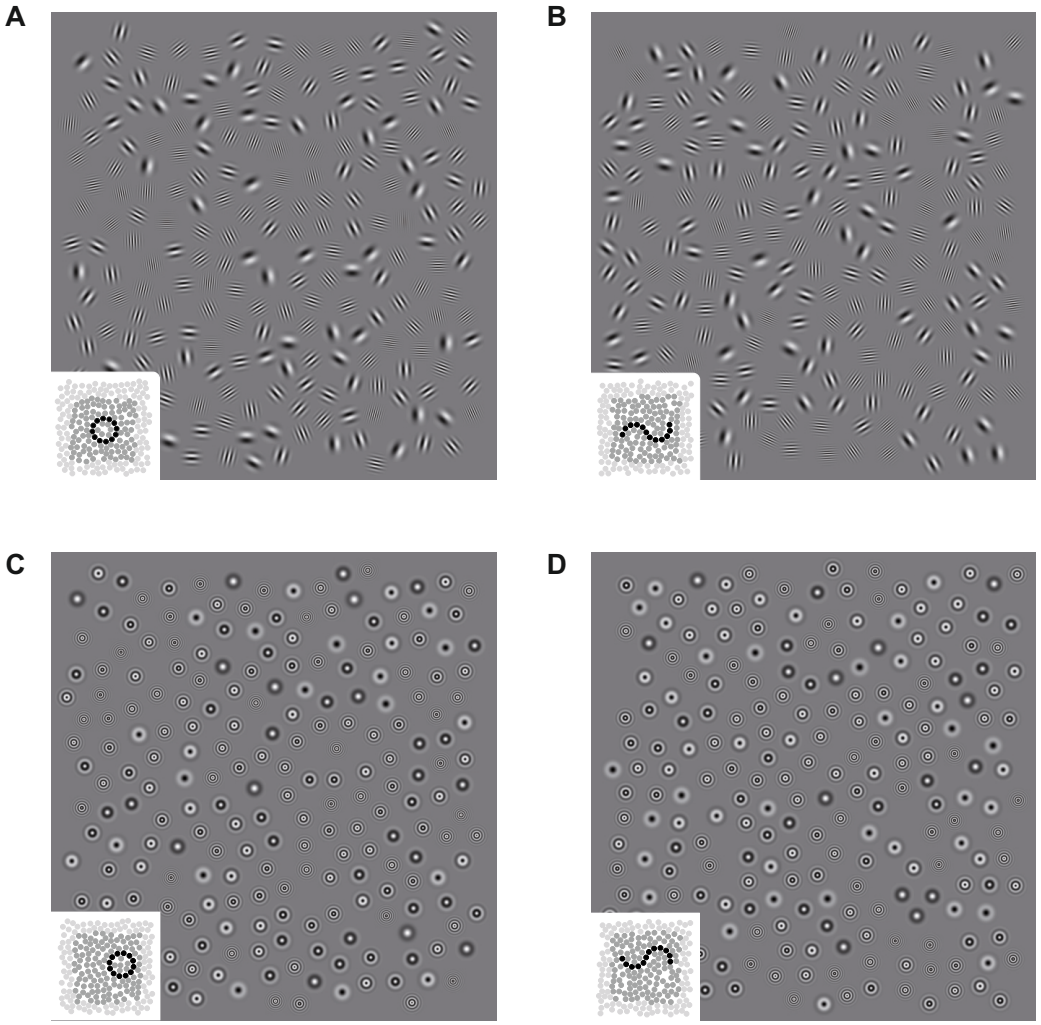

Figure A.3: Example of target display stimuli for (Persike et al., 2009) Experiment Two. The figures contained in the boxes on the bottom left corner of each panel indicate the position and shape of the contour in the larger stimulus display. Left panels O-Shape contours, right panels S-Shape contours. Top two panels display stimuli generated with Gabor elements and bottom two panels display stimuli generated with radial stimuli. The spatial frequency of contour elements is homogeneous and has a value equivalent to the mean spatial frequency of background elements. Image obtained from (Persike et al., 2009).

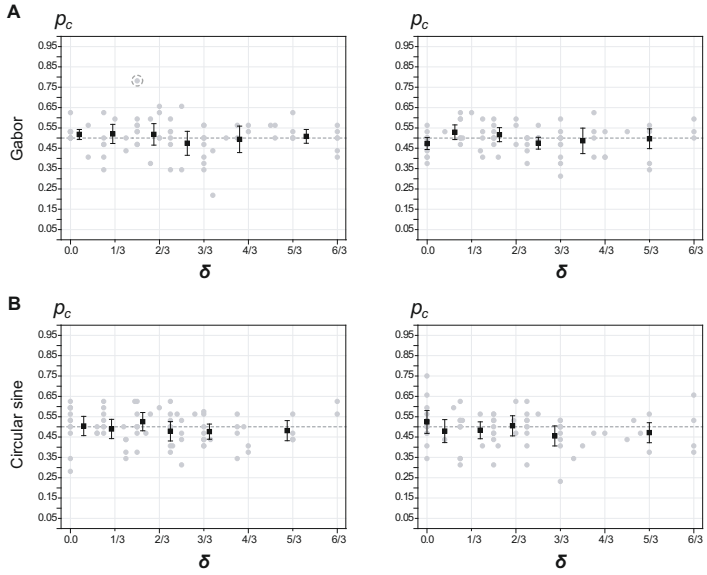

Figure A.4: Results from (Persike et al., 2009) of contour integration performance (in proportion correct) as a function of contour spatial frequency jitter homogeneity δ. Black squares in each display show the average proportion correct with a 95% confidence interval. Panels **A**, and **C** O-Shape stimulus results. Panels **B** and **D** S-Shape stimulus results. Upper row corresponds to stimuli with Gabor patches as elements in the display and the lower row corresponds to stimuli with circular elements in the display. An outlier is marked in panel **A** by circling the data point with a dashed line. Image obtained from (Persike et al., 2009).

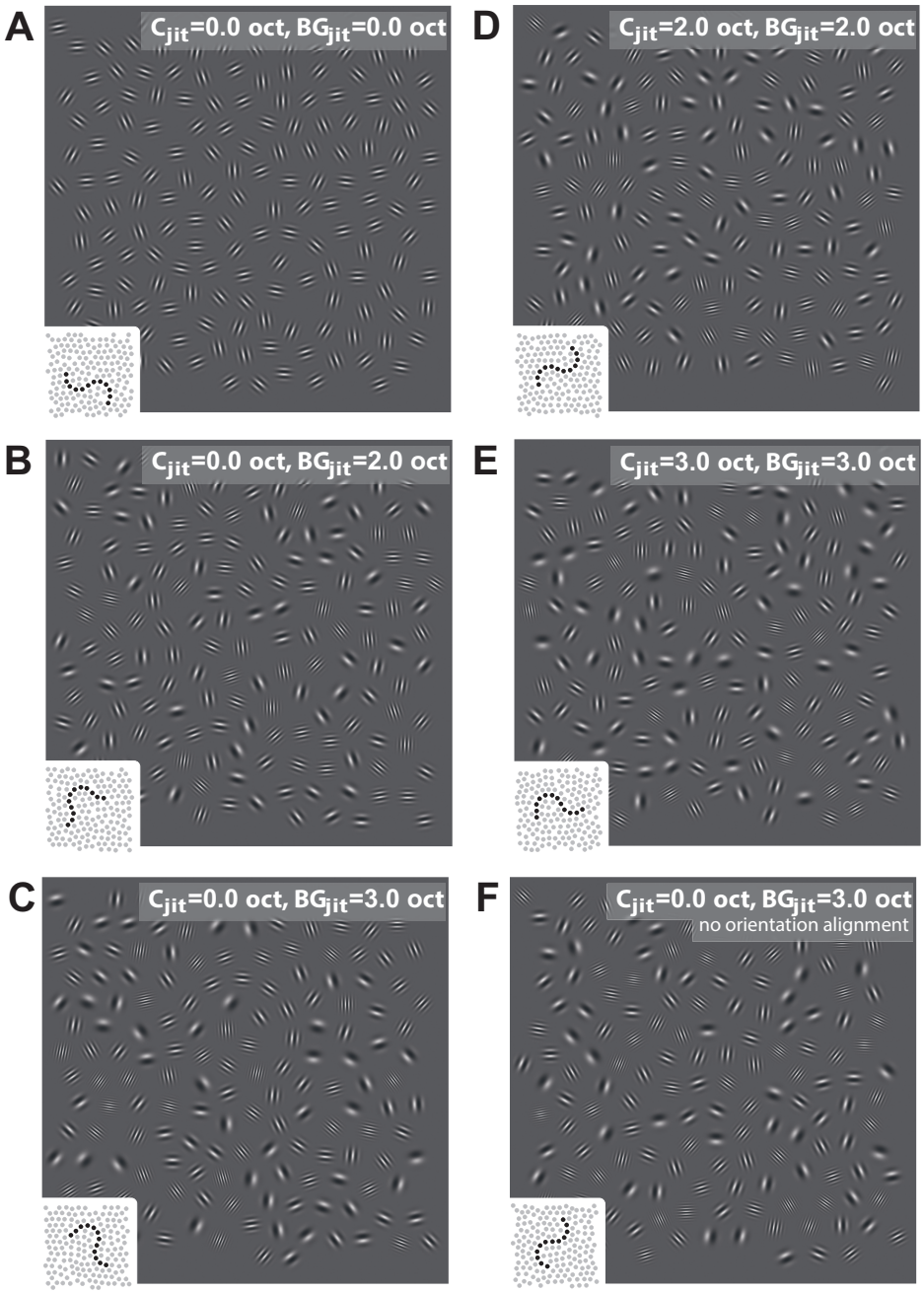

Figure A.5: Example of target display stimuli for (Persike and Meinhardt, 2015b). The stimuli consisted of aligned contours (with the exception of panel **F**) with spatial frequencies for elements in the display as indicated in each of the panels. The figures contained in the boxes on the bottom left corner of each panel indicate the position and shape of the contour in the larger stimulus display. Image obtained from (Persike and Meinhardt, 2015b).

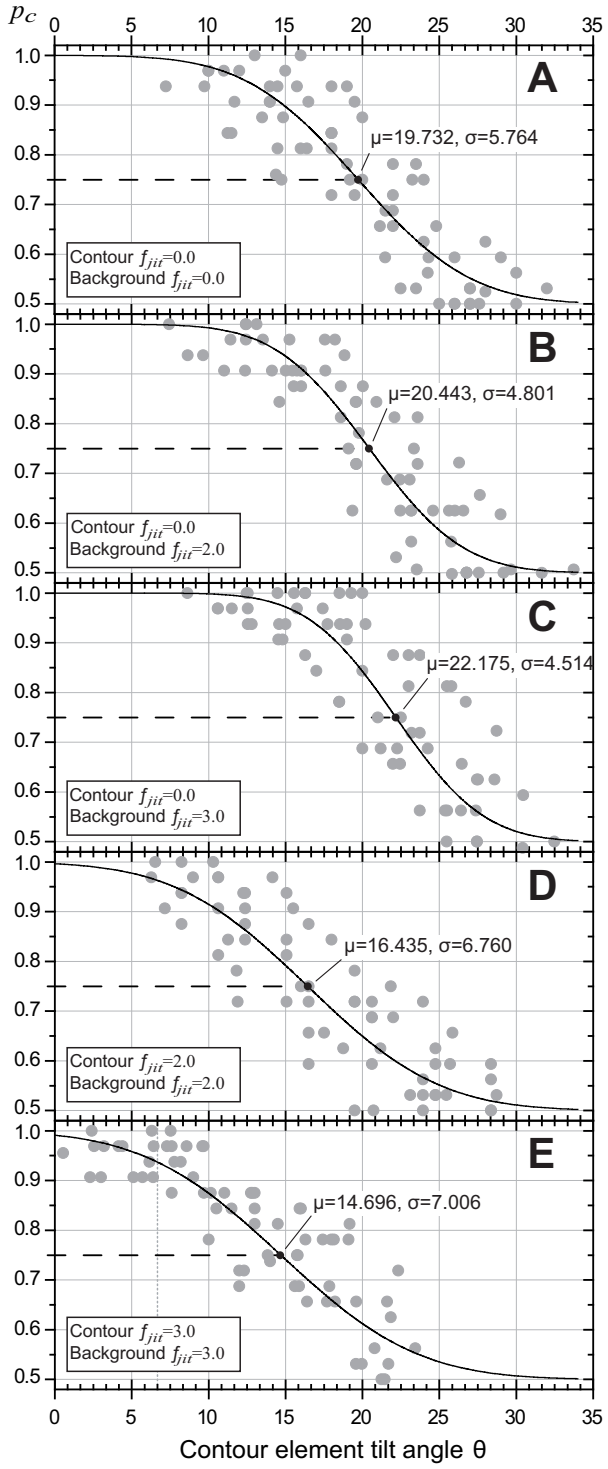

Figure A.6: Psychometric functions for the conditions depicted in the different panels of figure A.5 (with the exception of panel **F** in figure A.5). Fits are depicted with a cumulative Gaussian function intersection at the between subject mean tilt angle threshold (μ) and with a mean standard deviation estimate (σ). The means of the 75% tilt angle threshold marked by a black circle. Image obtained from (Persike and Meinhardt, 2015b).

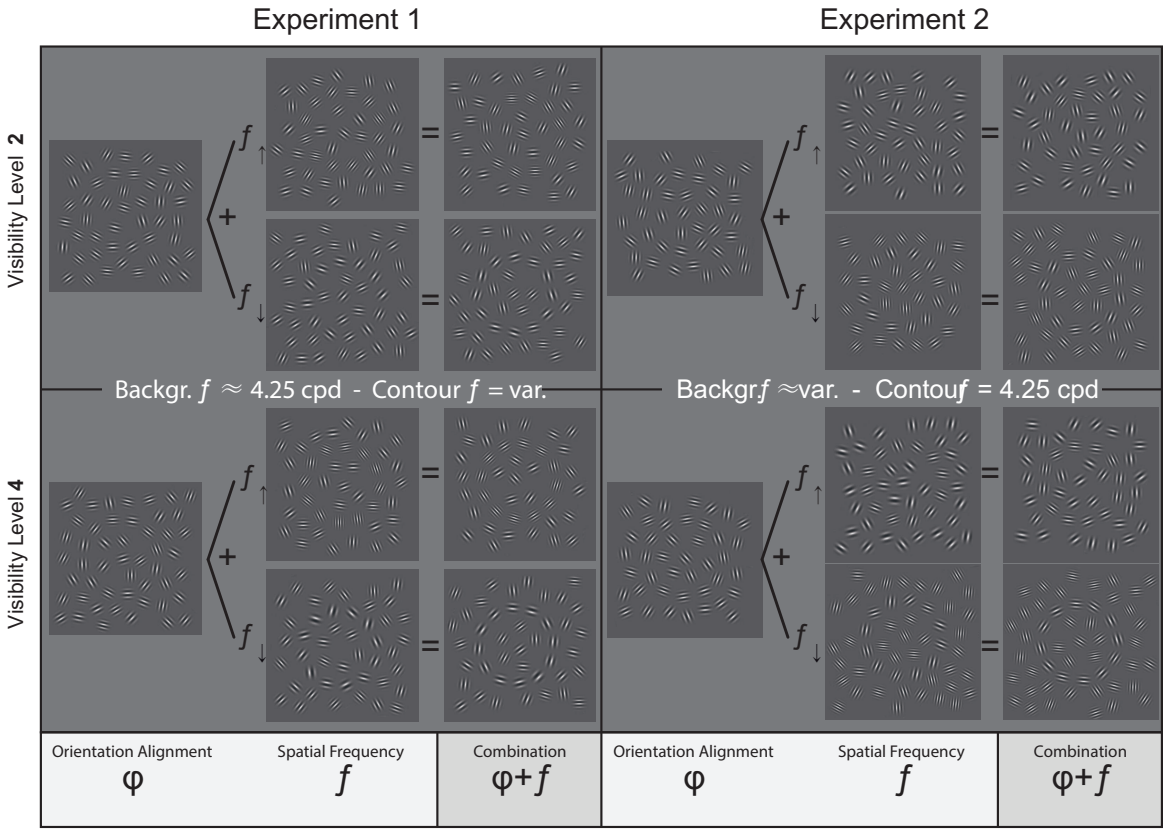

Figure A.7: Example of target display stimuli for Persike and Meinhardt (2015a). Visibility levels 2 and 4 are shown, as indicated. The orientation cues are shown alone on the left most display in each of the panels, then an example of spatial frequency shifts in the upwards and downwards direction are shown, and finally the combination of the two. Stimuli for Experiment One (spatial frequency shift on contour elements), and Experiment Two (spatial frequency shift on background elements) are shown as indicated in the figure. Image obtained from (Persike and Meinhardt, 2015a).

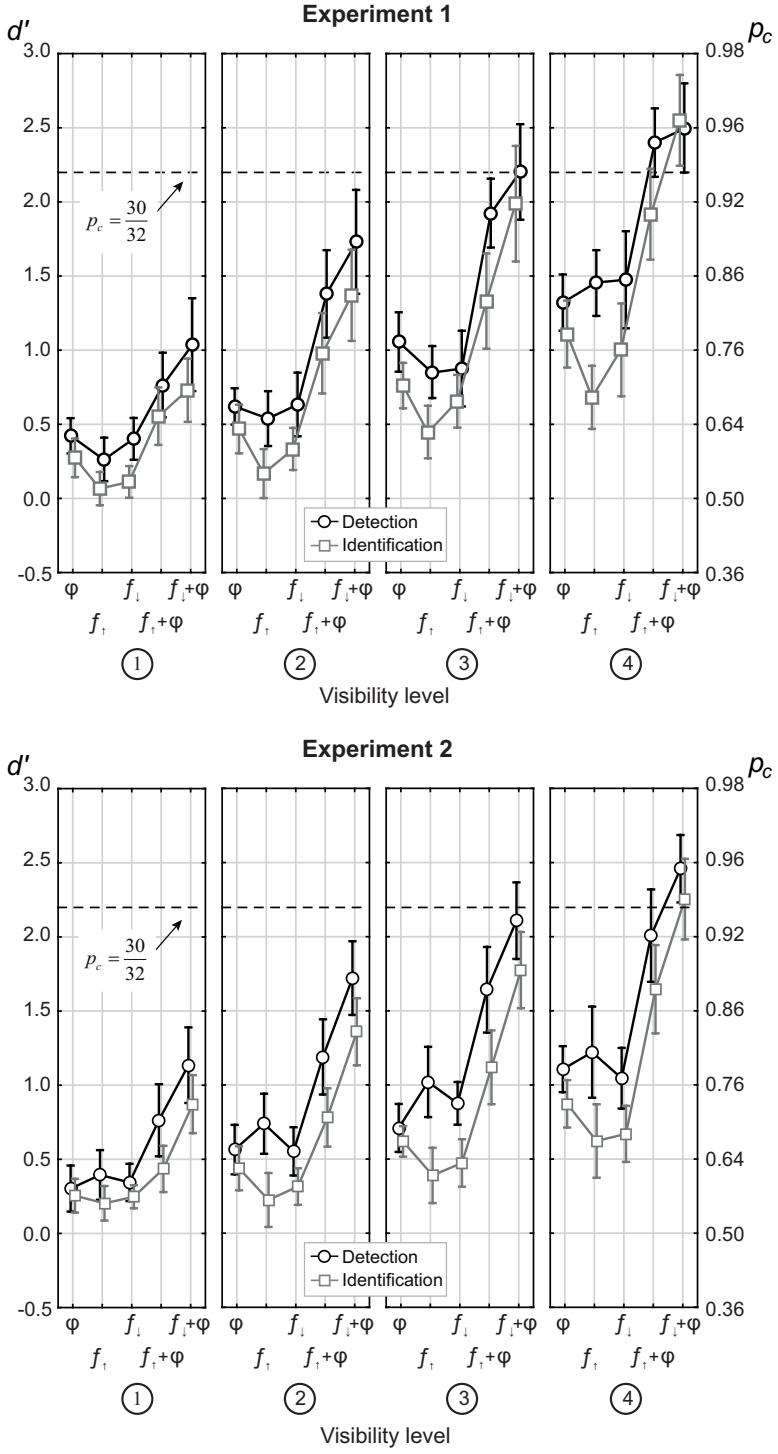

Figure A.8: Data for Experiments One and Two of Persike and Meinhardt (2015a). Mean d' and mean proportion correct for contours defined by orientation only (ϕ), an upwards or downwards (indicated by the arrows) spatial frequency shift (f), or a combination of an upwards or downwards spatial frequency shift and orientation. Four different visibility levels for baseline performance of contour detection with a single feature as indicated in the figures. Contour detection performance is shown in black and identification performance is shown in grey. Image obtained from (Persike and Meinhardt, 2015a).

Appendix B

B.1 Spatial Frequency Couplings

During the developing phase of the model the couplings given by the spatial frequency differences between the edge elements in E were not realized as described in equation 4.8. During this phase the coupling were realized as a discount factor W^{f_d} assumed to monotonously depend on the absolute spatial frequency difference $\Delta f_{ij} := |log_2 f_i - log_2 f_j|$ between two line segments. W^{f_d} was bounded by 1 for $\Delta f_{ij} = 0$ and by 0 for $\Delta f_{ij} \to \infty$.

Since the exact functional dependency between W^{f_d} and Δf can be arbitrary, a dimensionless ("effective") spatial frequency, γ (replacing f) was introduced, so that $f_i \to \gamma_i$. Consistently, spatial frequency differences were defined as $\Delta\gamma_{ij} := |\gamma_i - \gamma_j|$. W^{f_d} was linearly related to $\Delta\gamma$ via $W^{f_d}(\Delta\gamma) = \max\{1 - \Delta\gamma, 0\}$, thus, rectifying all values of $\Delta\gamma$ larger than 1.

Spatial frequency shifts and jitters applied to either contour elements, or background elements in E (with respect to f_0) were expressed in terms of γ. If a frequency shift of $\Delta\gamma$ was applied to the background elements in E, contour elements had a γ of f_0, while background elements had a γ value of $\Delta\gamma$. Similarly, if a spatial frequency jitter of size σ_γ was applied to either background elements, contour elements, or all elements in E, the values for the γs of the corresponding edge elements were drawn from a uniform distribution between $\pm\sigma_\gamma/2$. For simplicity purposes during this phase $f_0 = 0$, thus $\gamma_0 = 0$.

B.1.1 From Discount Factors to True Spatial Frequency Couplings

Evidence for long range interactions of neurons tuned to similar spatial frequencies (in the visual cortex) suggests that the interactions between these neurons are better described by an exponential decay rather than by a linear function (Boucsein et al., 2011). Thus, the simplification used during the developing phase of our model, and described above, is likely not biophysically plausible. In order to make the model biophysically plausible by adapting the monotonic function used during the developing

phase of the model to an exponential decay function the following procedure was carried out:

1. A number of parameters were selected which **(A)** had a good performance in the orientation only case[1] (i.e.: the model was required to outperform the humans when taking into account an error window computed with binomial statistics[2]); and **(B)** performance in the SF only condition (i.e.: shifts experiment - see section 4.2) reached at least visibility level 4, or higher, also while accounting for measurement errors predicted by binomial statistics (visibility level 4 was 80% performance, see section 4.2 or Persike and Meinhardt (2015a)).

2. On this selection of parameters a fit was performed in which γ (the discount factor used to represent spatial frequency differences) was to be matched to the physical measure used in the experiments (i.e.: octaves).

   In order to do this fit first the visibility levels (V1 through V4 used in Persike and Meinhardt (2015a)) were mapped onto the orientation only condition of the model. If at a tilt angle of 19° we found a 75% performance for humans, we can also determine which performance we find for the model in the orientation only condition at a tilt angle of 19° (which was usually higher). With this procedure performance levels for humans ($\boldsymbol{P}_{H(V1)}...\boldsymbol{P}_{H(V4)}$) which correspond to V1 through V4 in (Persike and Meinhardt, 2015a) were found, and performance levels for the model ($\boldsymbol{P}_{M(V1)}...\boldsymbol{P}_{M(V4)}$) which corresponds to $\boldsymbol{P}_{H(V1)}...\boldsymbol{P}_{H(V4)}$ were found - see panel **A** in figure B.1). In addition $\boldsymbol{P}_{M75\%}$ which corresponds to $\boldsymbol{P}_{H75\%}$ was also found.

   A performance curve for the model in the spatial frequency shifts condition was also obtained, and from the measures $\boldsymbol{P}_{H(V1)}...\boldsymbol{P}_{H(V4)}$ and $\boldsymbol{P}_{M(V1)}...\boldsymbol{P}_{M(V4)}$ a range $[Z_{\boldsymbol{P}_{H(V1...V4)}}$ to $Z_{\boldsymbol{P}_{M(V1...V4)}}]$ of γs which yielded a performance between $\boldsymbol{P}_{H(V1)}...\boldsymbol{P}_{H(V4)}$ and $\boldsymbol{P}_{M(V1)}...\boldsymbol{P}_{M(V4)}$ was obtained (see panel **B** in figure B.1).

   Finally, a correct-performance matrix in dependence of tilt angle and spatial frequency jitter was generated for the manipulation levels of both variables (this was akin to (Persike and Meinhardt, 2015b)), and a fit was done on the data and evaluated on a finely sampled variable space - see panel **C** in figure B.1. In this plot two contour lines are present, they represent $\boldsymbol{P}_{M75\%}$ and $\boldsymbol{P}_{H75\%}$. The range between these two contour lines at tilt angles 16.4° and 14.6° was obtained and dubbed $[Z_{\boldsymbol{P}_{H(2Oct,3Oct)}}$ to $Z_{\boldsymbol{P}_{M(2Oct,3Oct)}}]$ since it represents the difference in performance between the model and subjects at the tilt angle thresholds found for a 75% correct performance (human) with a 2 and 3 octave jitter on spatial frequency. Note that when calculating the ranges $[Z_{\boldsymbol{P}_{H(2Oct,3Oct)}}$ to $Z_{\boldsymbol{P}_{M(2Oct,3Oct)}}])$ it is important to keep in mind that absolute differences computed from pairs of samples drawn from a random distribution with a certain width ψ will have an average difference much smaller than ψ (for a uniform distribution, for example, this value is $\psi/3$).

[1] See section 4.2 for a description of the different test scenarios - i.e.: cases.

[2] See Appendix C.1 for a description of how the confidence interval.

3. With the ranges ($[Z_{\boldsymbol{P}_{H(V1...V4)}}$ to $Z_{\boldsymbol{P}_{M(V1...V4)}}]$ and $[Z_{\boldsymbol{P}_{H(2Oct,3Oct)}}$ to $Z_{\boldsymbol{P}_{M(2Oct,3Oct)}}]$) obtained in the previous steps a dependency between the physical measure used in the psychophysical experiments (i.e.: octaves) and the γ was created and plotted, and a function was fitted through these points (see panel **D** in figure B.1).

After having gone through the described process for a large number of parameter sets a decaying exponential was found to be a good fit between the discount factor γ and the physical units used in the psychophysical experiments (i.e.: octaves).

When doing the final selection of parameters which presented a good fit a final constrain was placed on the data to ensure that a good area in the parameter space was found. It was ensured that there was a relationship in the spatial frequency jitter conditions in which performance decreased with increasing jitter on all elements in the display, and also where performance stayed relatively constant with different levels of SF jitter on background elements only.

Note that the fit and parameter selection was done based on the *Fixed Scaling Model* results, this had some implications in the qualitative match of the *Variable Scaling Model* results.

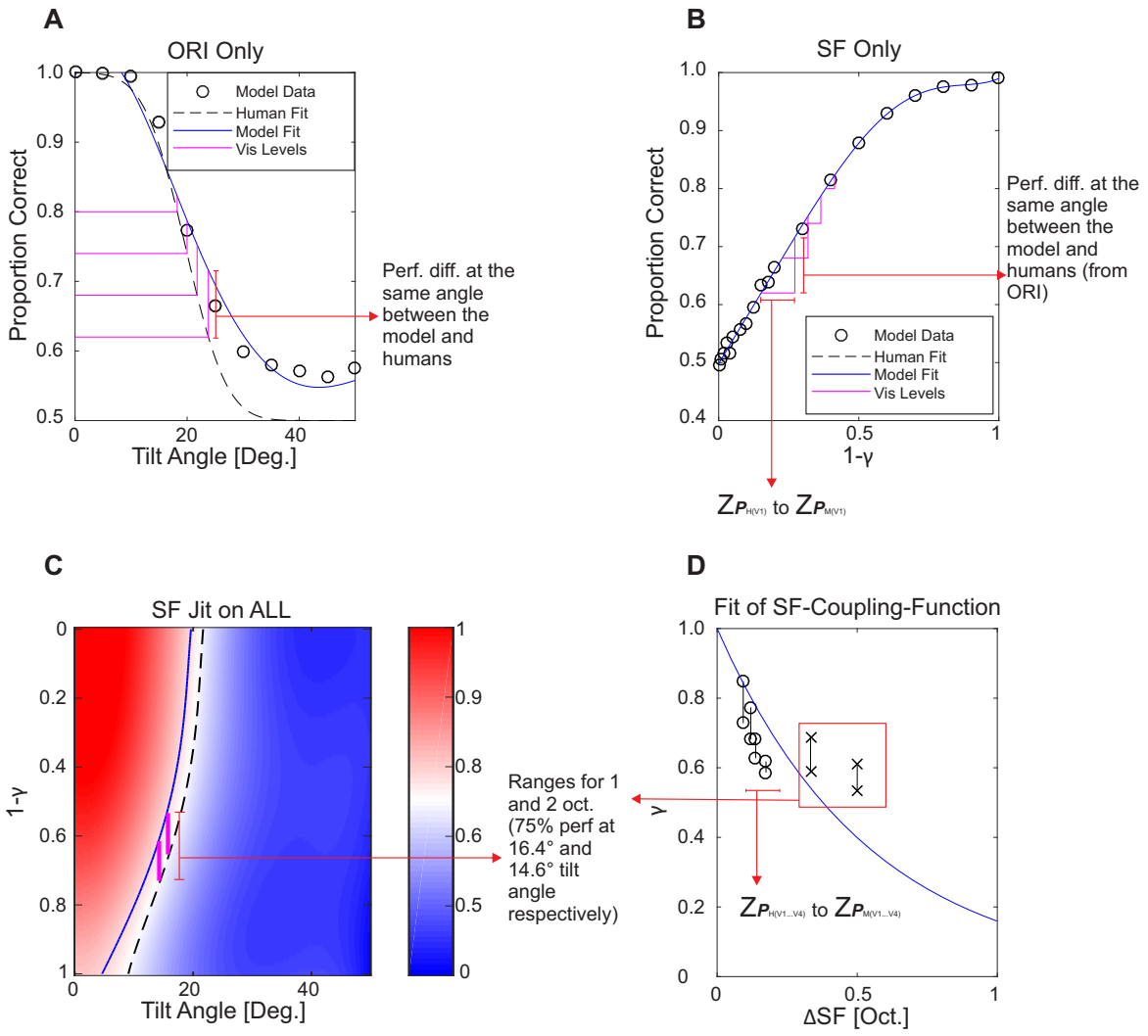

Figure B.1: **A**: Psychometric curve of the orientation only condition. Circles: model data; Dashed line: human fit; Solid blue line: Model Fit; Horizontal magenta lines mark performance at the four visibility levels of Persike and Meinhardt (2015b) on the human fit curve. Vertical magenta lines mark the performance at the human fit curve and at the model data fit curve for the angle at which the corresponding visibility level was found on the human fit curve. **B**: Psychometric curve of the spatial frequency shift condition only with $1-\gamma$ as the independent variable. Vertical magenta lines indicate the performance difference between the human fit and the model fit found for the 4 visibility levels in the orientation only condition, the horizontal (magenta) lines indicate ranges in γ corresponding to the performance differences. **C**: Correct-performance matrix in dependence of tilt angle and spatial frequency jitter. Solid blue line: contour line at $\boldsymbol{P}_{M75\%}$ proportion correct, Black dashed line: contour line at $\boldsymbol{P}_{H75\%}$ proportion correct. Magenta lines indicate the range in $1-\gamma$ at 16.4° and 14.6° of tilt angle between $\boldsymbol{P}_{H75\%}$ correct and $\boldsymbol{P}_{M75\%}$ correct. **D**: Fit of γ to octaves. The fit attempts to intersect all six ranges marked, a decaying exponential was used to fit the data. All arrows and text boxes are put in place for clarification purposes, please see the relevant sections of text for further clarification.

Appendix C

C.1 Binomial Statistics

To calculate the confidence intervals on the results obtained from the model, binomial statistics were employed. A table of lower and upper confidence interval bounds was found for a discretization of performance from 0 to 1 proportion correct, in N steps[1]. The discretization took the form of $P_j = j/N...N/N$.

A binary search was performed to calculate the lower bound of the confidence interval. The end of the binary search was reached when the difference between S_{lb} and S_{ub} was smaller than Δ^{sig}, a predefined precision value determining how close S_{lb} should be to the real value of the lower bound the of the confidence interval, for any P_j. S_{lb} and S_{ub} were the lower, and upper bounds (respectively), of the search range in the binary search.

The binomial cumulative distribution function

$$B(P_j, x) = \sum_{i=0}^{x} \binom{N}{i} P_j^i (1 - P_j)^{(N-i)} \tag{C.1}$$

was employed to find $B(P_j)$, the probability of observing up to x successes in N independent trials, where the probability of success in any given trial was P_j.

The binary search was initialized with $S_{lb} = B_{j-1}$ for $j > 0$. If $j = 0$ then $S_{lb} = 0$. For the upper bound in the binary search S_{ub} was initialized as $S_{ub} = x/N$. Throughout the binary search S_{lb} or S_{ub} were replaced by $P^{mid} = (S_{lb} + S_{ub})/2$. S_{lb} was replaced by P^{mid} when $C^{mid} \leq \Delta^{sig}$, otherwise S_{ub} was replaced by P^{mid}; and $C^{mid} = B(P^{mid})$.

Once the binary search ended for any P_j the lower bound of the confidence interval C_j^{low} was set to S_{lb}. Once all lower bounds were found for all P_js in the discretization, the upper bound C^{high} was generated by reversing the vector C^{low} and taking its complementary probability.

Having a table with three vectors, one with lower bounds of the confidence interval (C_j^{low}), one with upper bounds of the confidence interval (C_j^{high}), and one with per-

[1] N was equal to the number of trials employed to get any given mean performance. See section 4.5.2.

formance discretized in N steps (P_j), allowed for an easy way to look-up a confidence interval for any P_j.

C.2 Selection of Parameters

In this appendix a detailed description is given of how the parameters for presentation were chosen.

Since in the first parameter search conducted, a large number of unique parameters could qualitatively reproduce the results (with several of the readout mechanisms), a selection of suitable parameters was made based on the criterion described below. In order to keep a parameter set in consideration for evaluation of further psychophysical results the parameter set was required to:

1. Deliver a performance for the **ORI** condition which was equal, or higher, than human performance at the same tilt angle. In order to account for error margins, the upper bound of the error predicted by binomial statistics for the model was required to have an equal, or better, performance than the lower error bound predicted for (the fit of) the psychophysical data (see figure 4.23).

2. Deliver a performance for the $\mathbf{SF}_{\uparrow}$ and $\mathbf{SF}_{\downarrow}$ conditions which was equal, or higher, than 80% correct (i.e.: visibility level 4 in Persike and Meinhardt (2015b)). The same procedure as described above was used to account error margins.

After applying this criterion, the number of parameter sets deemed suitable was reduced to 123, from the initial 648 unique parameter sets selected from the infinite parameter space. 36 unique parameters were suitable for the bidirectional model, and 87 for the unidirectional model. On these 123 parameter sets the procedure described in Appendix B.1.1 was applied in order to evaluate the model results with the same physical units as in the psychophysical experiments.

After the model results were evaluated with physical units in the **SF** conditions (rather than with the discount factor - see Appendix B.1.1) a further reduction of suitable parameters was done, reducing the suitable parameters from 123 to 10. Of these 10 parameter sets, 3 were suitable for the bidirectional model and 7 for the unidirectional model. This reduction was done based on the goodness of the fit done on the discount factor γ (see Appendix B.1.1), on a visual inspection of the **SF** condition, and the three spatial frequency jitter conditions (jitter on all elements, on background elements only, or on contour elements only). For a comprehensive description of how parameter spaces were selected for exploration, the selections criteria employed, and how a narrowing of the parameter space was conducted to select the parameters used for presentation, please refer to the flow chart below.

Stage One: ORI at 15°

Steps Taken:
Test **ORI** only condition was tested at 15° of tilt angle. All possible combinations of the following values for each of the individual parameters were tested:

- λ_{ex}: [0.25, 0.4, 0.65, 1]
- λ_{in}: [0.25, 0.4, 0.65, 1]
- σ_{α}: [0.1:0.1:0.6]
- σ_{β}: [0.2:0.2.1.2]
- W_0^{ex}: [4, 6, 8, 10, 15, 20, 30, 40]
- W_0^{in}: [0.2, 0.5, 1, 1.5, 2, 3, 5, 10]

• **Condition for further consideration:** Perform roughly between 75-90%. Manual check.

↓

General conclusions:

Unidirectional model performs better than bidirectional model, however, both lead to the same insights:

- W_0^{in} in intermediate rage yield the best results → continue with W_0^{in}: [1...3]
- No particular region for W_0^{ex} looks promissing → continue with full range
- Best results with intermediate spatial scales (λ_{ex} and λ_{in}), λ_{in} should be equal or larger than λ_{ex} → continue with fixed values for λ_{ex} and λ_{in}
- Very good results when $\sigma_{\beta}>2*\sigma_{\alpha}$
- Estimators: k_{sum}, k_{square}, s^2 deliver good results; k_{max}, delivers OK results; s^3, s^4 does not deliver good results

Stage Two: (a) Full ORI and SF Psycho. curves; (b) Full field for tilt angle and SF Jitters

Steps Taken:
A. Test full psychometric curve for tilt angle and spatial frequency shift
B. Test full field for tilt angle and spatial frequency jitters (SF Jitter conditions: Background Only, Contour and Background, Contour Only). Full field means that all tilt angles were tested with all possible jitter levels.

Parameters tested:

- λ_{ex}: 0.4
- λ_{in}: 0.4
- σ_{α}: [0.1:0.1:0.6]
- σ_{β}: [0.2:0.2.1.2]
- W_0^{ex}: [4, 6, 8, 10, 15, 20, 30, 40]
- W_0^{in}: [1, 1.5, 2, 3]

• **Condition for further consideration:**
1. Deliver a psychometric curve for tilt angle and SF Shift conditions (i.e.: not flat performance). Manual check.
2. Deliver the following patters for SF Jitter conditions: (a) Increase perf or stay relatively stable for increasing levels of background jitter; (b) Decrease performance with increasing levels of contour and background jitter; (c) Deliver high performance regardless of jitter on contour. **Of most relevance was (b)**. Manual check.

↓

General conclusions:

- Mostly in the range of W_0^{ex} = 10 to 30 and W_0^{in} = 1 to 2 do we find the desired pattern of results
- σ_{β} should be larger than σ_{α}; pairs in the ranges of σ_{α}= 0.1 to 0.3 and σ_{β}=0.6 to 0.8 deliver good results
- With this restricted range now new pairs of λ_{ex} and λ_{in} can be tried

Stage Three: Constraints in performance & on Fit

Steps Taken:

A. Test full psychometric curve for tilt angle and spatial frequency shift
B. Test full field for tilt angle and spatial frequency jitters (SF Jitter conditions: Background Only, Contour and Background, Contour Only)
C. Fit discount factor (γ) to octaves. See "From Discount Factors to True Spatial Frequency Couplings - Section 4.4.3.1"

Parameters tested: All possible combinations of

- σ_α: [0.1, 0.2, 0.3]
- σ_β: [0.6, 0.8]
- W_0^{ex}: [10, 30]
- W_0^{in}: [1, 1.5, 2]
- $[\lambda_{ex}, \lambda_{in}]$: [*0.4 0.4; 0.5 0.5; 0.4 0.6; 0.5 0.6*]

- **Condition for further consideration:**
  1. Must perform at the same level or above humans (accounting for confidence intervals). Automatized test.
  2. Must achieve at least 80% or more correct detection in SF shift condition. Automatized test.
  3. Must show a decrease in performance for the jitter on contour and background condition as jitter levels increase. Manual check.
  4. Exponential fit should pass (roughly) through the six ranges extracted (see section "From Discount Factors to True Spatial Frequency Couplings" - 4.4.3.1). Manual check.

General conclusions:

- An exponential decay function delivers a good fit for the discount factor into octaves.
- Five clusters in regions of the parameter space which deliver good performance appear.

Cluster 1:
-Three values for W_0^{in}: 1, 1.25, 1.5
-Two values for $[\sigma_\alpha, \sigma_\beta]$ pairs: [0.1, 0.6], [0.6, 0.8]
-Fixed value for W_0^{ex}: 10
-Fixed value for λ_{ex}: 0.4
-Fixed value for λ_{in}: 0.6
Six possible combinations

Cluster 2:
-Three values for W_0^{in}: 1.5, 1.8, 2
-Two values for $[\sigma_\alpha, \sigma_\beta]$ pairs: [0.2, 0.6], [0.2, 0.8]
-Fixed value for W_0^{ex}: 10
-Fixed value for λ_{ex}: 0.4
-Fixed value for λ_{in}: 0.6
Six possible combinations

Cluster 3:
-Two values for $[\sigma_\alpha, \sigma_\beta]$ pairs: [0.2, 0.6], [0.2, 0.8]
-Two values for λ_{ex}: 0.4, 0.5
-Fixed for W_0^{in}: 1
-Fixed value for W_0^{ex}: 30
-Fixed value for λ_{in}: 0.6
Four possible combinations

Cluster 4:
-Three values for W_0^{in}: 1.5, 1.8, 2
-Two value for λ_{ex}: 0.4, 0.5
-Fixed values for $[\sigma_\alpha, \sigma_\beta]$: [0.3, 0.8]
-Fixed value for W_0^{ex}: 30
-Fixed value for λ_{in}: 0.6
Six possible combinations

Cluster 5: Fixed combinations!

$[\sigma_\alpha, \sigma_\beta]$:	[0.1, 0.8]	[0.1, 0.8]	[0.1, 0.8]	[0.1, 0.6]	[0.2, 0.8]
W_0^{in}:	1	1.25	1.5	1.5	1
W_0^{ex}:	10	10	10	10	10
λ_{in}:	0.6	0.6	0.6	0.5	0.5
λ_{ex}:	0.5	0.5	0.5	0.4	0.4

Five fixed combinations

Stage Four: Test model with physical units

Steps Taken:
A. Test full psychometric curve for tilt angle and spatial frequency shift
B. Test full field for tilt angle and spatial frequency jitters (SF Jitter conditions: Background Only, Contour and Background, Contour Only).
C. Look for λ_f
Spatial frequency manipulations applied with spatial frequency rather than discount factor

Parameters tested: 5 regions in space yielding **27** distinct parameters combinations. See General conclusions, Stage Three.

• **Condition for further consideration:** λ_f which yields results for all visibility levels in the SF shift conditions and in the Jitter conditions must exist. Manual check.

General conclusions:

- Ten parameter sets fulfill the λ_f requirements

Stage Five: Decide on parameters for presentation

Steps Taken:
A. Test full psychometric curve for tilt angle and spatial frequency shift
B. Test full field for tilt angle and spatial frequency jitters (SF Jitter conditions: Background Only, Contour and Background, Contour Only).
Spatial frequency manipulations applied with spatial frequency rather than discount factor and a selected λ_f

Parameters tested: 10 distinct parameter sets, 7 for the Unidirectional model and 3 for the Bidirectional model

• **Condition for further consideration:** yields the best compromise between all results in all conditions. Checked manually

General conclusions:

- Two parameter sets selected for presentation, one for unidirectional mode, one for bidirectional mode

Unidirectional:

$[\sigma_\alpha, \sigma_\beta]$ pairs: [0.2, 0.8]
W_0^{ex}: 30
W_0^{in}: 1.8
λ_{ex}: 0.5
λ_{in}: 0.6
λ_f: 1.45

Bidirectional:

$[\sigma_\alpha, \sigma_\beta]$ pairs: [0.1, 0.8]
W_0^{ex}: 10
W_0^{in}: 1
λ_{ex}: 0.5
λ_{in}: 0.6
λ_f: 1.15