



KATHOLIEKE UNIVERSITEIT LEUVEN

Faculteit Psychologie en Pedagogische Wetenschappen

Laboratorium voor Experimentele Psychologie

Contextual modulation in low-level
visual processing:
Psychophysical, neurophysiological and
computational perspectives

Proefschrift aangeboden
tot het verkrijgen van de graad van
Doctor in de Psychologie

door

Nathalie Van Humbeeck

Promotor: Prof. Dr. Johan Wagemans

2016

Nathalie Van Humbeeck

Contextual modulation in low-level visual processing: psychophysical, neurophysiological and computational perspectives

Dissertation submitted to obtain the degree of Doctor in Psychology, June 2016

Supervisor: Prof. Dr. Johan Wagemans

The response to a sensory stimulus depends heavily on its spatial and temporal context. A wide number of studies have investigated contextual effects of perceptual and neurophysiological responses. These studies primarily focused on the primary visual cortex (V1), reporting that responses of V1 classical receptive fields are influenced by the surrounding context. One type of observed contextual modulation in V1 is collinear facilitation, which entails the enhancement of neural responses to an element surrounded by collinear elements. Such facilitatory interactions are thought to form the neural basis of contour perception. The dependence of these interactions on the relative orientation and position of elements has been described in terms of an association field.

A number of computational models have used the association field concept to predict human performance in detecting a contour embedded in a background of randomly oriented elements. These models typically construct a saliency map of the image indicating which elements are likely to be part of a contour. Their validity can be assessed by evaluating how well they can predict the image regions that will attract eye movements during contour integration. In the first study, observers' eye movements were recorded during a contour integration task. An association field model was able to predict saccade targets. In addition, we showed that fixation duration and saccade size followed a time course which depended on the saliency and percept of a contour.

The results of the first study showed that the presence of potential contours in the image influenced observers' eye movements during contour integration. This suggests that an initial, but still incomplete, saliency map actively guided eye movements. However, it remained unclear which brain processes were involved in the presaccadic selection of a saccade target. In a second study, we aimed to examine whether presaccadic EEG activity is modulated by the presence of a contour or a region with high association strength in peripheral vision. We found that presaccadic EEG activity, mainly over parietal and occipital brain areas, was predictive of the distance between the saccade landing position and the contour. In addition, when a contour was absent, presaccadic activity predicted the association strength at the saccade landing position. Our results suggest that the presaccadic amplitude reflects the degree to which top-down processes can override bottom-up saliency.

Other spatio-temporal contextual effects on low-level visual processing have been studied in the context of apparent motion (AM). AM refers to the percept of motion occurring when two stationary stimuli are alternately presented at two different locations. It has been found that the detectability of stimuli is reduced in the presence of AM. Previous studies have attributed such masking to interference caused by AM-induced excitation, claiming that V1 neurons respond as if a stimulus is physically moving. In a third study, we investigated this claim by modelling grating detectability during AM using a physiologically inspired population code model. The model predicted only a small amount of V1 activation, which could not account for the observed masking nor for any perceptual completion of the motion path. Our model revealed that AM masking is instead due to strong suppression of V1 responses, which is consistent with the theoretical framework of predictive coding.

A fourth study provided further evidence in favor of AM-induced suppression by measuring and modelling perceived grating contrast during AM. Multiple studies have found that the perceived contrast of a grating is reduced when V1 responses to that grating are suppressed. If AM indeed suppresses neural responses to gratings at the level of V1, the perceived contrast of the grating should be reduced in the presence of AM. A population code model similar to the model presented in the third study provided a full account of performance in a contrast discrimination task. The model indeed revealed a reduction in perceived contrast caused by strong AM-induced suppression. A model only incorporating AM-induced excitation could not account for the data.

Nathalie Van Humbeeck

Contextuele modulatie in low-level visuele verwerking: een psychofysische, neurofisiologische en computationele benadering

Proefschrift aangeboden tot het verkrijgen van de graad van Doctor in de Psychologie, juni 2016
Promotor: Prof. Dr. Johan Wagemans

De respons op een sensorische stimulus hangt sterk af van zijn spatiale en temporele context. Een groot aantal studies hebben contextuele effecten van perceptuele en neurofisiologische responsen onderzocht. Deze studies focusten vooral op de primaire visuele cortex (V1) en rapporteerden dat responsen van V1 klassieke receptieve velden beïnvloed worden door de omringende context. Eén vorm van geobserveerde contextuele modulatie in V1 is collineaire facilitatie, wat verwijst naar de versterking van neurale responsen op een element omgeven door collineaire elementen. Zulke faciliterende interacties worden verondersteld de neurale basis te vormen van contourperceptie. De afhankelijkheid van deze interacties van de relatieve oriëntatie en positie van elementen is beschreven door middel van een associatieveld.

Een aantal computationele modellen hebben het associatieveldconcept gebruikt om menselijke performantie te voorspellen in het detecteren van een contour ingebed in een achtergrond van random georiënteerde elementen. Deze modellen construeren een saliëntiemap van het beeld dat aangeeft welke elementen waarschijnlijk tot een contour behoren. Hun validiteit kan worden beoordeeld door te kijken hoe goed ze de regio's in het beeld voorspellen die oogbewegingen aantrekken tijdens contourintegratie. In de eerste studie werden de oogbewegingen van observeerders geregistreerd tijdens contourintegratie. Een associatieveldmodel kon de fixatielocaties voorspellen. We toonden ook aan dat fixatieduur en de grootte van saccades een tijdsverloop volgden dat afhing van de saliëntie en perceptie van een contour.

De resultaten van de eerste studie toonden aan dat de aanwezigheid van potentiële contours in het beeld oogbewegingen van observeerders beïnvloedde tijdens contourintegratie. Dit suggerert dat een initiële, maar nog incomplete, saliëntiemap actief oogbewegingen leidde. Het bleef echter onduidelijk welke hersenprocessen betrokken waren in de presaccadische selectie van een fixatiepositie. In een tweede studie beoogden we te onderzoeken of presaccadische EEG activiteit gemoduleerd wordt door de aanwezigheid van een contour of een regio met hoge associatiesterkte in het perifere zicht. We vonden dat presaccadische EEG activiteit, vooral over pariëtale en occipitale hersengebieden, voorspellend was voor de afstand tussen het eindpunt van de saccade en de contour. Wanneer geen contour aanwezig was, voorspelde presaccadische activiteit de associatiesterkte aan het saccadische eindpunt. Onze resultaten suggereren dat de presaccadische EEG amplitude de mate reflecteert waarin top-down processen bottom-up saliëntie kunnen overschrijven.

Andere spatio-temporele contextuele effecten op low-level visuele verwerking werden bestudeerd in de context van apparaante beweging (AB). AB verwijst naar het bewegingspercept dat optreedt wanneer twee stationaire stimuli afwisselend gepresenteerd worden op twee verschillende locaties. Men heeft aangetoond dat AB de detecteerbaarheid van stimuli reduceert. Eerdere studies hebben zulke masking toegewezen aan interferentie veroorzaakt door AB-geïnduceerde excitatie, claimend dat V1 neuronen reageren alsof een stimulus fysiek beweegt. In een derde studie hebben we deze bewering onderzocht door middel van het modelleren van detecteerbaarheid van sinusgolfpatronen tijdens AB, gebruikmakend van een fysiologisch geïnspireerd populatiecoderingsmodel. Het model voorspelde enkel een kleine hoeveelheid V1 activatie, welke te klein was om de geobserveerde maskering of perceptuele invulling langs het bewegingspad te verklaren. Ons model toonde aan dat AB maskering het gevolg is van sterke suppressie van V1 responsen, wat consistent is met het theoretische kader van predictive coding.

Een vierde studie leverde aanvullende evidentiële voor AB-ingeïnduceerde suppressie door middel van het meten en modelleren van gepercipieerd contrast tijdens AB. Meerdere studies hebben gerapporteerd dat suppressie van V1 responsen op een sinusgolfpatroon het gepercipieerd contrast van het sinusgolfpatroon reduceert. Als AB inderdaad V1 responsen onderdrukt, dan zou gepercipieerd contrast van sinusgolfpatronen tijdens AB moeten gereduceerd worden. Een populatiecoderingsmodel, gelijkend op het model uit de derde studie, vatte de performantie in een contrast discriminatietaak volledig. Het model toonde inderdaad een reductie in gepercipieerd contrast aan veroorzaakt door AB-geïnduceerde suppressie. Een model dat enkel gebruik maakt van AB-geïnduceerde excitatie kon de data niet voorspellen.

Dankwoord

Het is nog moeilijk om te bevatten, maar nu ben ik dan toch aan het einde van mijn doctoraat gekomen. Ik besef hoe uniek deze periode is geweest. Vele momenten zullen me nog lang blijven en verschillende mensen hebben daarin een belangrijke rol gespeeld.

Allereerst zou ik mijn promotor Johan Wagemans willen bedanken. Johan, bedankt om mij de kans te geven een doctoraat te starten. Door jou kwam ik in een heel boeiende en stimulerende onderzoeksomgeving terecht. Bedankt ook voor de vrijheid die je me hebt gegeven om mijn eigen weg te gaan en me daarin te steunen. Thank you Udo, for your help during my first study, and for your valuable advice and enthusiasm. The visits at your lab in Bremen were always very inspiring and enjoyable. Frouke wil ik bedanken omdat ik altijd bij haar terecht kon tijdens mijn eerste jaar doctoraat. I would also like to thank Andrey, for his time, guidance and support.

Agna, bedankt voor je hulp op administratief vlak, maar ook voor je luisterend oor. Ook wil ik Hanne bedanken, het was zeer fijn om jou als bureaugenoot te hebben: we deelden over veel dingen dezelfde mening en ik kon bij jou ook al eens mijn frustraties kwijt. Bedankt Sander, je was al in het lab toen ik begon en je bent er nog steeds. Voor mij was dat zo'n beetje een geruststellende gedachte, dat er ook dingen zijn die niet zo snel veranderen. Bedankt Lore, om bij te dragen tot de gezellige sfeer op het bureau en aan de lunchtafel. Nihan, you always seemed to know when there was something I wanted to vent about: thank you for being such a great listener. Ruth, bedankt voor de bezoekjes aan onze bureau en ons te komen opvrolijken als je vond dat we te bedrukt keken. Pieter, het was altijd fijn om met jou te babbelen over wetenschap en andere dingen, bedankt daarvoor. Karolien, Katrijn, Leen en Ilse, jullie wil ik zeker bedanken om recepties en feestjes na de werkuren soms zeer memorabel te maken. Het was de laatste jaren toch niet helemaal hetzelfde meer zonder jullie. Bedankt Francis en Dries, en in het laatste jaar ook Maja en Sofie, het was bijzonder fijn om samen met jullie de statistiekpractica te verzorgen.

Een heleboel andere collega's wil ik ook bedanken om bij te dragen tot een gezellige sfeer in het lab (deze lijst is zeker niet volledig): Mark, Bart, Caitlin, Rudy, Helene, Leia, Einat, Sofia, Radha, Birgitt, Maarten, Vebjörn, Kris, Steven, Sam, en Kathleen.

Laura, Ine, en Katie, mijn beste vriendinnen voor het leven, bedankt dat ik altijd op jullie kan rekenen. Bedankt ook om op mijn werk steeds voor de nodige afleiding te zorgen en om me op de hoogte te houden van de laatste nieuwtjes uit Boechout. Ilse, ik ben echt blij dat we er nog altijd in slagen om geregeld af te spreken. Bedankt voor je vriendschap, ik weet dat ik altijd bij jou terecht kan.

Sandy, bedankt om mijn zus te zijn, je steunt me soms meer dan je zelf beseft. Op vele vlakken lijken we op elkaar en daardoor zijn er heel wat dingen die enkel jij en ik begrijpen. Mijn ouders zou ik ook willen bedanken voor hun onvoorwaardelijke steun en vertrouwen, en voor de kansen die ik heb gekregen om verder te studeren.

Er zijn zo van die momenten die de rest van je leven volledig op zijn kop kunnen zetten. Het moment dat ik jouw bureau binnenstapte, was er zeker zo één. Tom, vanaf die eerste dag wist ik al wie je was en dat is eigenlijk nooit veranderd. Ik weet wat jij in je mars

hebt en bewonder je daarvoor enorm. Bedankt voor alle mooie momenten die we al samen hebben meegeemaakt. Bedankt ook om steeds de tijd en energie te vinden om mij te helpen en advies te geven. En tenslotte is er Elias, mijn lief klein wondertje. Er is niets fascinerender op deze wereld dan jou zien opgroeien. Bedankt om zelfs de meest stresserende dag op te vrolijken met jouw schattige glimlach.

Nathalie

Contents

1 General Introduction	13
1.1 Contextual modulations in low-level visual processing	14
1.2 Collinear facilitation and the association field	15
1.3 Contextual influences of AM	18
1.4 Predictive coding theory	22
References	24
2 The role of eye movements in a contour detection task	29
2.1 Introduction	30
2.2 Methods	33
2.2.1 Participants	33
2.2.2 Apparatus	34
2.2.3 Stimuli	34
2.2.4 Procedure	35
2.2.5 Eye movement analysis	37
2.2.6 Architecture of contour integration model	38
2.2.7 Model dynamics	41
2.2.8 Fitting procedure	41
2.2.9 Prediction of fixation locations	42
2.3 Results	43
2.3.1 Task performance	43
2.3.2 Eye movement behavior	44
2.3.3 Prediction of fixation locations	48
2.4 Discussion	53
2.4.1 Effect of contour saliency	54
2.4.2 Dynamics of eye movements	56
2.4.3 Predictive power of contour integration model	58
2.4.4 Differences between observers	60
2.4.5 Saccades compensate for reduced saliency in peripheral vision . .	61
2.4.6 Conclusions	61
References	63
3 Presaccadic EEG activity reflects visual saliency during contour integration	67
3.1 Introduction	68
3.2 Methods	73

3.2.1	Participants	73
3.2.2	Stimuli	73
3.2.3	Procedure	75
3.2.4	Eye movement recording	75
3.2.5	EEG recording	76
3.2.6	Eye movement analysis	77
3.2.7	EEG preprocessing	77
3.2.8	Generalized additive mixed modelling	78
3.3	Results	84
3.3.1	Response and eye movement behavior	84
3.3.2	Results for contour-present trials	84
	Random-effect variables Participants and Image	84
	Distance between saccade landing position and contour	85
	Effects of eye movement parameters	86
3.3.3	Results for contour-absent trials	88
	Random-effect variables Participants and Image	88
	Association strength at the saccade landing position	89
	Effects of eye movement parameters	90
3.3.4	Variance explained by GAMM	92
3.4	Discussion	93
3.5	Appendix A: Grand average and random effects for contour-present trials	98
3.6	Appendix B: Grand average and random effects for contour-absent trials	99
	References	100
4	Apparent motion suppresses responses in V1: a population code model	105
4.1	Introduction	106
4.2	Results	108
4.2.1	AM masking	108
4.2.2	Population code model of AM-induced effects	110
4.2.3	Evaluation of model fit	113
4.2.4	Response suppression during AM	114
4.2.5	AM-induced effects are tuned to orientation	115
4.2.6	Early excitation is not the cause of perceptual filling-in	116
4.3	Discussion	117
4.4	Materials and Methods	121
4.4.1	Subjects	121
4.4.2	Apparatus	122
4.4.3	Stimuli	122
4.4.4	Detection task	122
4.4.5	Population code model	124
4.4.6	Model constraints and fitting	130
4.5	Appendix A: Individual data and pooling	131
	References	135
5	Response suppression during apparent motion reduces perceived grat- ing contrast	139

5.1	Introduction	140
5.2	Materials and Methods	141
5.2.1	Subjects	141
5.2.2	Apparatus	141
5.2.3	Stimuli	141
5.2.4	Discrimination task	142
5.2.5	Population code model	143
	Equations	143
	Deriving relative matching contrast	146
	Model constraints and fitting	146
5.3	Results	147
5.3.1	AM biases choice proportions	147
5.3.2	AM reduces perceived grating contrast	148
5.3.3	Perceived grating contrast reduction is caused by response suppression	150
5.3.4	The role of AM-induced excitation	154
5.3.5	Orientation tuning of AM-induced suppression	157
5.3.6	Comparison with the population code model of Chapter 4	159
5.4	Discussion	161
5.5	Appendix A: Inter-subject variability	168
	References	171
6	General Discussion and Conclusions	173
6.1	Summary of results	174
6.2	Bottom-up saliency and eye movements	176
6.3	Predictive coding	179
6.4	Conclusion	182
	References	183

Chapter 1

General Introduction

As human beings, it is crucial for our survival to react to changes in our spatial environment and light is an important messenger that enables us to do that. Changes in our environment are not intrinsically meaningful: they become meaningful through our interaction with them. In order to be able to adapt to our environment, it is crucial that the responses to previous interactions with our environment are stored, which will be activated when the organism encounters similar environmental changes in the future (Friston, 2011; Gibson, 1979). Hence, being able to “see” something can be interpreted as the act of being able to predict which interactions with the external environment are possible at a given moment. The idea that the visual system continuously updates predictions about the external world following sensory stimulation has also been put forward by predictive coding theories (Clark, 2013; Friston, 2005; Kok & de Lange, 2015; Rao & Ballard, 1999). These theories oppose the traditional view that visual information is processed by a strictly feedforward hierarchical system in which sensory information is passively processed at increasingly larger spatial scales and complexity as one ascends the visual hierarchy. According to predictive coding, brain regions at different levels in the hierarchy constantly communicate with each other about what they expect the other brain region is responding to. In this doctoral dissertation, we will mainly focus on the visual processing of local features at lower-levels of the visual hierarchy, examining how neural responses to these features are influenced by their spatial and temporal context when considering a hierarchical and predictive brain.

1.1 Contextual modulations in low-level visual processing

Visual information projected onto our retina is transmitted via retinal ganglion cells through the lateral geniculate nucleus (LGN) to the primary visual cortex (V1). Neurons in LGN and V1 have been shown to respond to only a small region of the retina or visual space. The region within which the presence of appropriate visual features will evoke an action potential in a particular neuron is defined as the classical receptive field (CRF) of that neuron. According to the definition of the CRF, stimulation outside the neuron’s CRF will thus not be sufficient to evoke a response. However, when the

neuron is excited by a stimulus presented inside its CRF, stimulation of the surrounds of the CRF can modulate the response of this neuron. This neural manifestation of contextual influences is called the extra-classical receptive field effect and has repeatedly been observed in V1. In most cases, the presence of the surround stimulus leads to a suppression of the spiking responses to the stimulus in the CRF (Jones, Grieve, Wang, & Sillito, 2001; Sengpiel, Sen, & Blakemore, 1997; Walker, Izumi, & Freeman, 2000). The contextual suppression is generally highest when the center and surround stimuli have similar orientations (i.e. iso-orientation suppression; Chao-Yi & Wu, 1994; DeAngelis, Freeman, & Ohzawa, 1994; Knierim & Van Essen, 1992; Levitt & Lund, 1997; Sillito, Grieve, Jones, Cudeiro, & Davis, 1995; Walker, Ohzawa, & Freeman, 1999). Likewise, maximal surround suppression has been observed when center and surround stimuli have similar spatial frequencies (DeAngelis et al., 1994; Chao-Yi & Wu, 1994; Walker et al., 1999) and similar speeds (Chao-Yi & Wu, 1994). In contrast, when the surround stimuli are orthogonal to the center stimulus, surround facilitation has been reported (Jones, Wang, & Sillito, 2002; Levitt & Lund, 1997; Sillito et al., 1995).

1.2 Collinear facilitation and the association field

Using bars or Gabor patches as surround stimuli, several studies have found facilitatory interactions, although these effects seem to be highly dependent on the contrast of the center stimulus (Levitt & Lund, 1997; Mizobe, Polat, Pettet, & Kasamatsu, 2001; Polat, Mizobe, Pettet, Kasamatsu, & Norcia, 1998; Toth, Rao, Kim, Somers, & Sur, 1996). Responses to a low-contrast stimulus inside the CRF are generally facilitated by the surrounding bar or Gabor path, while responses to a high-contrast stimulus are suppressed. More importantly, maximal facilitation has been found when the center and surround stimuli are collinear, which decreases when the distance between the center and surround stimuli increases (Chen, Kasamatsu, Polat, & Norcia, 2001; Kapadia, Ito, Gilbert, & Westheimer, 1995; Kapadia, Westheimer, & Gilbert, 2000; Knierim & Van Essen, 1992; W. Li, Piéch, & Gilbert, 2006; Mizobe et al., 2001; Nelson & Frost, 1985; Nothdurft, Gallant, & van Essen, 1999; Polat et al., 1998). It has been suggested that these collinear facilitation effects can be characterized by an “association field” mechanism, which was

first introduced by Field, Hayes, and Hess (1993). An association field describes a network of interactions between neighbouring oriented line segments depending on their relative orientations and spatial positions. Facilitatory interactions are strengthened with decreasing distance, curvature and deviation from co-circularity between the line segments. The shape of the association field has been found to closely resemble the co-occurrence statistics of edges in natural images (Elder & Goldberg, 2002; Geisler, Perry, Super, & Gallogly, 2001), providing support that the association field function is a mechanism evolved to optimize the interaction with the natural environment.

A network of long-range facilitatory V1 connections described by an association field has been proposed to form the neural basis of contour integration, although it is likely that these connections are supported by top-down feedback from higher-level areas involved in shape detection, perceptual learning and attention (W. Li, Piëch, & Gilbert, 2008). Contour integration is the process by which elongated contours consisting of co-aligned elements are extracted from cluttered images, which is a crucial step for further scene segmentation and object recognition. Psychophysical experiments on contour integration have typically used a stimulus in which a contour consisting of collinear elements (often Gabor patches) is embedded in a background of randomly oriented but otherwise similar elements. Local density cues in the image are absent so that observers can only detect the contour based on the orientation relationships between the elements. Observers' performance to detect such a contour is highly dependent on the average curvature or path angle of the contour, which is the average change in orientation between adjacent contour elements. Observers' ability to detect a contour is best when the contour forms a straight line and decreases as the degree of curvature of the contour increases (Field et al., 1993; Hess & Dakin, 1997; Watt, Ledgeway, & Dakin, 2008). Several computational models have implemented an association field concept to explain how the visual system extracts collinear contours from images (Ernst et al., 2012; Ernst, Mandon, Pawelzik, & Kreiter, 2004; Z. Li, 1998; Ullman, Gregory, & Atkinson, 1992; Yen & Finkel, 1998). Typically, an association field model computes an association strength value for each oriented element in the image which determines how likely it is that the element belongs to a contour (Watt et al., 2008).

The aforementioned contextual modulations in V1 (i.e., iso-orientation suppression and collinear facilitation) have been proposed to contribute to the bottom-up *saliency* of elements across an image (Z. Li, 2002; Spratling, 2012). In most models, bottom-up salient stimuli refer to stimuli which are intrinsically conspicuous because they differ from their surroundings with respect to a certain feature (Itti & Koch, 2001). These stimuli are assumed to attract attention and eye movements automatically and involuntarily. The spatial pattern of saliency values across an image is called a saliency map. For instance, in the models of Koch and Ullman (1985) and Itti and Koch (2001), a saliency map results from the integration of individual feature maps which are based on a calculation of local feature contrasts, i.e., differences in simple features (such as color or orientation) between a target and neighboring elements. Studies have shown that the distribution of saliency levels in a bottom-up saliency map based on feature contrast can predict, to a certain extent, the spatial distribution of eye movements during free-viewing of visual scenes. Observers tended to make more saccades to locations with high levels of feature contrast (DeCarlo & Santella, 2002; Parkhurst, Law, & Niebur, 2002). However, the proportion of fixations that can be accounted for by such a saliency map model is typically modest and fixation locations are known to be influenced by top-down factors such as task relevance and semantics of the scene (Henderson, 2003; Zelinsky, 2008). In the V1 saliency map model of Z. Li (2002), the saliency values of locations are determined by the responses of output neurons in V1. For instance, due to iso-orientation suppression, elements surrounded by similar orientations will be less salient than elements that differ in orientation. Similarly, an oriented element forming a smooth contour with its neighboring elements will be more salient due to collinear facilitation. To account for the latter surround effect, a neural connection structure is proposed whose shape closely resembles the aforementioned association field (see Z. Li, 1998 for more details). Feedback from higher-level visual areas serves to selectively modulate the sensitivity to the contour as a whole. This V1 model can account for a wide range of psychophysical experiments on contour integration, texture segmentation and visual search (Koene & Zhaoping, 2007; Z. Li, 1999; Zhaoping & Snowden, 2006; Zhaoping & May, 2007)¹.

¹Please note that Li Z. and Zhaoping L. refer to the same author Li Zhaoping.

Z. Li (2002) thus predicts that collinear elements provide salient locations in a bottom-up manner, before a global contour is detected. An association field model should thus be able to predict the locations at which observers are likely to make eye fixations during a contour integration task. This hypothesis was tested in our first study (**Chapter 2**). We examined observers' eye movement behavior during two free-viewing contour search tasks in which we manipulated task difficulty by varying the curvature of the contour. It is known that the detectability of a contour highly depends on how far it is located in the periphery, especially for curved contours. Making eye movements can thus considerably support contour integration by bringing possible contour candidates into the foveal region. Based on the association field model of Ernst et al. (2012), we constructed saliency maps of the images which were used in the two search tasks to determine whether observers fixated salient locations predicted by the model.

The results of our first study indicated that observers' eye movements were not at all random during contour search, but were affected to some degree by information about the saliency and position of a possible target contour. The aim of the second study (**Chapter 3**) was to explore the brain processes involved in saccade planning during a contour integration task. We tested the hypothesis that EEG activity in the interval prior to a saccade event is modulated by the presence of a contour or a region with high association strength in peripheral vision. Our results suggest that this presaccadic activity is indicative of top-down processes involved in overriding bottom-up saliency.

1.3 Contextual influences of AM

Psychophysical evidence indicates that the presence of a collinear context improves the detection of a target (Polat & Sagi, 1993; Popple, Polat, & Bonneh, 2001; Kapadia et al., 1995; Tzvetanov & Dresp, 2002; Zenger & Sagi, 1996). However, in a grating contrast discrimination task, Wu and Chen (2010) found that the presence of collinear flankers improved contrast discrimination performance at low grating contrasts, but impaired performance at high contrasts. The performance improvement was mainly the result of excitation, i.e., an increase in the low-level contrast response to a grating. The perfor-

mance impairment on the other hand was predominantly caused by low-level response inhibition. The mere presentation of flanker stimuli thus seems to cause a complex pattern of contextual modulation involving both excitatory and inhibitory effects. An interesting question is whether a similar complex pattern emerges for other, related instances of contextual modulation. One such instance, investigated by numerous previous studies, is apparent motion (AM). AM refers to the perception of a single moving stimulus when in fact two stationary stimuli (i.e., the AM-inducing stimuli) are presented in succession at two different spatial locations (Wertheimer, 1912). It has frequently been reported that the detection of a target stimulus presented along the perceived motion path is impaired, a phenomenon referred to as AM masking (Chong, Hong, & Shim, 2014; Hidaka, Nagai, Sekuler, Bennett, & Gyoba, 2011; Hidaka, Teramoto, & Nagai, 2012; Hogendoorn, Carlson, & Verstraten, 2008; Schwiedrzik, Alink, Kohler, Singer, & Muckli, 2007; Souto & Johnston, 2012; Yantis & Nakama, 1998). These findings have been explained by postulating that the activation of an internal representation of AM interferes with other visual stimuli physically presented on the motion path (Hidaka et al., 2011; Souto & Johnston, 2012; Yantis & Nakama, 1998). Hidaka et al. (2011) observed impaired detectability of a target Gabor grating presented along the AM trajectory and found that this impaired detectability disappeared with increasing orientation difference between the target and the AM-inducing stimuli. The authors interpreted their findings as evidence for perceptual “filling-in” of the motion path at lower-level areas, presumably V1. AM induces activation of neurons tuned to locations along the motion path, which interferes with the responses to Gabor stimuli physically presented along this path.

However, previous studies have never specified exactly what this interference entails. In **Chapter 4**, we claim that the exact nature of the AM-induced interference can be fully specified. If AM masking is indeed caused by perceptual filling-in at the level of V1, then AM masking should be considered equivalent to traditional pattern masking, in which a grating is masked by a simultaneously present background grating. Perceptual filling-in would increase V1 responses and thereby “interfere” much like a background grating would. The psychophysical and neural effects of the presence of a background grating on visual processing and perception have been widely studied and are well known. Estab-

lishing the link between pattern masking and AM therefore allows us to make specific predictions about the effects of perceptual filling-in and whether or not interference is to be expected. Indeed, the presence of a background grating in the context of pattern masking does not always cause interference. The degree of masking of the target grating depends on the contrast of the background: high background contrasts lower grating detectability considerably, while low background contrasts facilitate grating detectability.

This can be easily understood by considering that grating detection seems to reflects responses of V1 neurons, which respond non-linearly to grating contrast (Heeger, 1992a, 1992b). To be able to detect a target grating on top of a background grating, a sufficiently large differential response is needed: the target grating presented in combination with the background grating should evoke a larger response at the level of V1 than the background grating alone. The larger this response difference between the target present and target absent case, the higher detection performance will be. It turns out that this response level increases when changing the background grating contrast from zero (no background grating present) to an intermediate contrast level. When background contrast is further increased, the response difference again decreases. This is a consequence of the non-linear response pattern of typical V1 neurons, as detailed in Figure 1.1. In the absence of a background grating, or at low background grating contrasts, a relatively high target grating contrast is required to obtain a sufficiently large differential response. The reason is that V1 contrast response functions are shallow at low contrasts. At intermediate contrasts, however, the functions are considerably steeper due to non-linear response expansion (Heeger, 1992a). When the background grating contrast is increased to an intermediate level, the contrast response function operates in this steep regime. Consequently, even a small increase of the target contrast on top of the background grating contrast causes a large response difference as small contrast perturbations are amplified by the relatively steep slope. For this reason, the presence of a background grating facilitates grating detection: detection performance is better compared to the situation in which the background is absent. This observation has been referred to as the pedestal effect (Nachmias & Sansbury, 1974). The steepness of the contrast response

function and, consequently, the differential response, decreases at high contrasts due to non-linear response compression in the context of contrast gain control (Heeger, 1992b). Correspondingly, at high background grating contrasts, facilitation disappears and the background grating indeed causes interference, i.e., masking. Only relatively high target contrasts on top of the background contrast will result in sufficiently large differential responses.

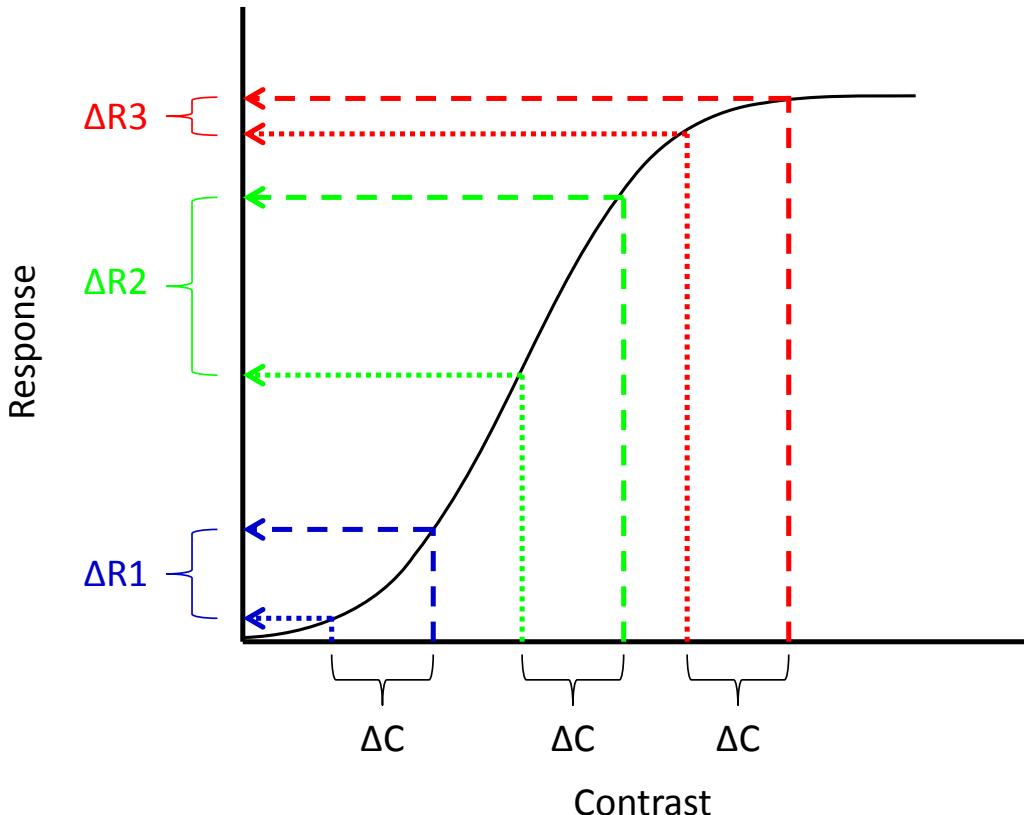


Figure 1.1. The full black line denotes a typical V1 contrast response function. The horizontal dotted lines mark the response to a background grating of low (blue), intermediate (green) and high (red) contrast. The corresponding dashed lines mark the response to a target grating, presented at a contrast level ΔC units higher than the background grating. The contrast difference ΔC between target and background grating is constant across background grating contrasts in this example. In a detection task in the context of a typical pattern masking experiment, observers have to detect this contrast difference based on the response difference ΔR . A small increase in target grating contrast relative to the background contrast leads to a much higher response difference for intermediate background contrast levels due to the large steepness of the contrast response function at these contrast levels. At low and high background grating contrasts, the contrast response function is much shallower and response differences are smaller.

In sum, a background grating can cause both facilitation and masking by evoking responses at the level of V1. Consequently, perceptual filling-in during AM does not neces-

sarily cause masking. Determining whether perceptual filling-in is a plausible explanation of AM-induced masking was the main goal of the study described in **Chapter 4**. We started by measuring grating detectability both in the absence and presence of AM for a wide range of target contrast levels and measured full psychometric functions. It is rather surprising that this has not yet been done by previous studies, which all measure detection performance at a single, arbitrarily-chosen target contrast level. This, of course, makes it particularly challenging to get a complete quantitative picture of the details of AM masking.

The psychometric function analysis revealed peculiar and hitherto unnoticed features of AM masking. To account for these features, we developed a physiologically-plausible population code model. This model, inspired by and highly similar to traditional pattern masking models, accounts for all data. Based on the model, we were able to arrive at surprising conclusions about the nature of AM-induced interference and, more generally, about the effects of AM on low-level neural representations. Our model also allowed us to formulate new predictions about the effect of AM on the appearance rather than the mere detectability of gratings. More specifically, our model predicts that AM should significantly alter the apparent contrast of gratings. In the study reported in **Chapter 5**, we test this prediction by determining whether and how AM affects perceived contrast.

1.4 Predictive coding theory

In this doctoral dissertation, we will propose that the results of **Chapter 4** and **5** can be explained within a predictive coding framework of sensory processing. Predictive coding provides a general computational framework describing how the brain processes sensory input (Friston, 2005; Mumford, 1992; Rao & Ballard, 1999). A key feature is that bottom-up sensory stimulation is combined with top-down expectations about the outside world to determine our perception. Predictions about sensory stimulation based on an internal model of the world are compared with “actual” sensory inputs to calculate the mismatch or residual error. Each brain area in the cortical visual hierarchy represents both these predictions and prediction errors. Different areas in the cortical hierarchy

continuously interact to find the hypothesis that best explains the input, thereby reducing the prediction error. Feedforward pathways carry the residual error between predicted and actual lower-level input to a higher-level area and trigger this area to update its hypothesis about the sensory input. Feedback connections from a higher-level cortical area then carry the updated prediction of the sensory input to a lower-level area. Via these feedback connections, the prediction error in lower-level areas is reduced when higher-level areas can accurately predict the sensory input. Different neural implementations of predictive coding differ in how exactly the feedback connections reduce the prediction error represented in lower-level areas, an issue which we will come back to in the General Discussion (**Chapter 6**).

Predictive coding theory thus predicts that local features which can be predicted from their surrounding spatio-temporal context will result in reduced activity in lower-level areas. Recent fMRI studies seem to support this view, showing that stimulus predictability suppresses responses in V1 (Alink, Schwiedrzik, Kohler, Singer, & Muckli, 2010; Den Ouden, Friston, Daw, McIntosh, & Stephan, 2009; Harrison, Stephan, Rees, & Friston, 2007; Fang, Kersten, & Murray, 2008; Murray, Kersten, Olshausen, Schrater, & Woods, 2002; Schellekens, van Wezel, Petridou, Ramsey, & Raemaekers, 2014). Murray et al. (2002) found increased activity in the lateral occipital complex, a higher-level cortical region critical for shape processing, and reduced activity in V1 when lines were arranged into a coherent 3D-shape. Another study by Alink et al. (2010) found that visual stimuli reduce responses in V1 when their onset or motion direction can be predicted from a surrounding AM context.

We will end this dissertation with a general discussion in **Chapter 6**, where we will attempt to derive some general conclusions from the presented work and interpret the results of our studies in light of predictive coding theory.

References

- Alink, A., Schwiedrzik, C. M., Kohler, A., Singer, W., & Muckli, L. (2010). Stimulus predictability reduces responses in primary visual cortex. *Journal of Neuroscience*, 30(8), 2960–2966.
- Chao-Yi, L., & Wu, L. (1994). Extensive integration field beyond the classical receptive field of cat's striate cortical neurons – classification and tuning properties. *Vision Research*, 34(18), 2337–2355.
- Chen, C.-C., Kasamatsu, T., Polat, U., & Norcia, A. M. (2001). Contrast response characteristics of long-range lateral interactions in cat striate cortex. *Neuroreport*, 12(4), 655–661.
- Chong, E., Hong, S. W., & Shim, W. M. (2014). Color updating on the apparent motion path. *Journal of Vision*, 14(14), 1–12.
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(03), 181–204.
- DeAngelis, G. C., Freeman, R. D., & Ohzawa, I. (1994). Length and width tuning of neurons in the cat's primary visual cortex. *Journal of Neurophysiology*, 71(1), 347–374.
- DeCarlo, D., & Santella, A. (2002). Stylization and abstraction of photographs. In *Acm transactions on graphics (tog)* (Vol. 21, pp. 769–776).
- Den Ouden, H. E., Friston, K. J., Daw, N. D., McIntosh, A. R., & Stephan, K. E. (2009). A dual role for prediction error in associative learning. *Cerebral Cortex*, 19(5), 1175–1185.
- Elder, J. H., & Goldberg, R. M. (2002). Ecological statistics of Gestalt laws for the perceptual organization of contours. *Journal of Vision*, 2(4), 324–353.
- Ernst, U. A., Mandon, S., Pawelzik, K. R., & Kreiter, A. K. (2004). How ideal do macaque monkeys integrate contours? *Neurocomputing*, 58, 971–977.
- Ernst, U. A., Mandon, S., Schinkel-Bielefeld, N., Neitzel, S. D., Kreiter, A. K., & Pawelzik, K. R. (2012). Optimality of human contour integration. *PLoS Computational Biology*, 8(5), e1002520.
- Fang, F., Kersten, D., & Murray, S. O. (2008). Perceptual grouping and inverse fMRI activity patterns in human visual cortex. *Journal of Vision*, 8(7), 2–9.
- Field, D. J., Hayes, A., & Hess, R. F. (1993). Contour integration by the human visual system: evidence for a local “association field”. *Vision Research*, 33(2), 173–193.
- Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1456), 815–836.
- Friston, K. (2011). The Implications of Embodiment: Cognition and Communication. In W. Tschauder & C. Bergomi (Eds.), (pp. 89–125). Exeter: Imprint Academic.
- Geisler, W. S., Perry, J. S., Super, B. J., & Gallogly, D. P. (2001). Edge co-occurrence in natural images predicts contour grouping performance. *Vision Research*, 41(6), 711–724.
- Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.
- Harrison, L., Stephan, K., Rees, G., & Friston, K. (2007). Extra-classical receptive field effects measured in striate cortex with fMRI. *Neuroimage*, 34(3), 1199–1208.

- Heeger, D. (1992a). Half-squaring in responses of cat striate cells. *Visual Neuroscience*, 9(5), 427–443.
- Heeger, D. (1992b). Normalization of cell responses in cat striate cortex. *Visual Neuroscience*, 9(2), 181–197.
- Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences*, 7(11), 498–504.
- Hess, R. F., & Dakin, S. C. (1997). Absence of contour linking in peripheral vision. *Nature*, 390(6660), 602–604.
- Hidaka, S., Nagai, M., Sekuler, A. B., Bennett, P. J., & Gyoba, J. (2011). Inhibition of target detection in apparent motion trajectory. *Journal of Vision*, 11(10), 1–12.
- Hidaka, S., Teramoto, W., & Nagai, M. (2012). Sound can enhance the suppression of visual target detection in apparent motion trajectory. *Vision Research*, 59, 25–33.
- Hogendoorn, H., Carlson, T. A., & Verstraten, F. A. (2008). Interpolation and extrapolation on the path of apparent motion. *Vision Research*, 48(7), 872–881.
- Itti, L., & Koch, C. (2001). Computational modeling of visual attention. *Nature Reviews Neuroscience*, 2(3), 194–203.
- Jones, H., Grieve, K., Wang, W., & Sillito, A. (2001). Surround suppression in primate V1. *Journal of Neurophysiology*, 86(4), 2011–2028.
- Jones, H., Wang, W., & Sillito, A. (2002). Spatial organization and magnitude of orientation contrast interactions in primate V1. *Journal of Neurophysiology*, 88(5), 2796–2808.
- Kapadia, M. K., Ito, M., Gilbert, C. D., & Westheimer, G. (1995). Improvement in visual sensitivity by changes in local context: parallel studies in human observers and in V1 of alert monkeys. *Neuron*, 15(4), 843–856.
- Kapadia, M. K., Westheimer, G., & Gilbert, C. D. (2000). Spatial distribution of contextual interactions in primary visual cortex and in visual perception. *Journal of Neurophysiology*, 84(4), 2048–2062.
- Knierim, J. J., & Van Essen, D. C. (1992). Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. *Journal of Neurophysiology*, 67(4), 961–980.
- Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiology*, 4(4), 219–227.
- Koene, A. R., & Zhaoping, L. (2007). Feature-specific interactions in salience from combined feature contrasts: Evidence for a bottom-up saliency map in V1. *Journal of Vision*, 7(7), 6–6.
- Kok, P., & de Lange, P. F. (2015). An introduction to model-based cognitive neuroscience. In U. B. Forstmann & E.-J. Wagenmakers (Eds.), (pp. 221–244). New York, NY: Springer, New York.
- Levitt, J. B., & Lund, J. S. (1997). Contrast dependence of contextual effects in primate visual cortex. *Nature*, 387(6628), 73–76.
- Li, W., Piëch, V., & Gilbert, C. D. (2006). Contour saliency in primary visual cortex. *Neuron*, 50(6), 951–962.
- Li, W., Piëch, V., & Gilbert, C. D. (2008). Learning to link visual contours. *Neuron*, 57(3), 442–451.
- Li, Z. (1998). A neural model of contour integration in the primary visual cortex. *Neural Computation*, 10(4), 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*, 96(18), 10530–10535.
- Li, Z. (2002). A saliency map in primary visual cortex. *Trends in Cognitive Sciences*, 6(1), 9–16.
- Mizobe, K., Polat, U., Pettet, M. W., & Kasamatsu, T. (2001). Facilitation and suppression of single striate-cell activity by spatially discrete pattern stimuli presented beyond the receptive field. *Visual Neuroscience*, 18(3), 377–391.
- Mumford, D. (1992). On the computational architecture of the neocortex. *Biological Cybernetics*, 66(3), 241–251.
- Murray, S. O., Kersten, D., Olshausen, B. A., Schrater, P., & Woods, D. L. (2002). Shape perception reduces activity in human primary visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 99(23), 15164–15169.
- Nachmias, J., & Sansbury, R. V. (1974). Grating contrast: discrimination may be better than detection. *Vision Research*, 14(10), 1039–1042.
- Nelson, J., & Frost, B. (1985). Intracortical facilitation among co-oriented, co-axially aligned simple cells in cat striate cortex. *Experimental Brain Research*, 61(1), 54–61.
- Nothdurft, H.-C., Gallant, J. L., & van Essen, D. C. (1999). Response modulation by texture surround in primate area V1: correlates of “popout” under anesthesia. *Visual Neuroscience*, 16(1), 15–34.
- Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research*, 42(1), 107–123.
- Polat, U., Mizobe, K., Pettet, M. W., Kasamatsu, T., & Norcia, A. M. (1998). Collinear stimuli regulate visual responses depending on cell’s contrast threshold. *Nature*, 391(6667), 580–584.
- Polat, U., & Sagi, D. (1993). Lateral interactions between spatial channels: suppression and facilitation revealed by lateral masking experiments. *Vision Research*, 33(7), 993–999.
- Popple, A., Polat, U., & Bonneh, Y. (2001). Collinear effects on 3-Gabor alignment as a function of spacing, orientation and detectability. *Spatial Vision*, 14(2), 139–150.
- Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2(1), 79–87.
- Schellekens, W., van Wezel, R. J., Petridou, N., Ramsey, N. F., & Raemaekers, M. (2014). Predictive coding for motion stimuli in human early visual cortex. *Brain Structure and Function*, 1–12.
- Schwiedrzik, C. M., Alink, A., Kohler, A., Singer, W., & Muckli, L. (2007). A spatio-temporal interaction on the apparent motion trace. *Vision Research*, 47(28), 3424–3433.
- Sengpiel, F., Sen, A., & Blakemore, C. (1997). Characteristics of surround inhibition in cat area 17. *Experimental Brain Research*, 116(2), 216–228.
- Sillito, A., Grieve, K., Jones, H., Cudeiro, J., & Davis, J. (1995). Visual cortical mechanisms detecting focal orientation discontinuities. *Nature*, 378(6556), 492–496.

- Souto, D., & Johnston, A. (2012). Masking and color inheritance along the apparent motion path. *Journal of Vision*, 12(7), 1–18.
- Spratling, M. W. (2012). Predictive coding as a model of the V1 saliency map hypothesis. *Neural Networks*, 26, 7–28.
- Toth, L. J., Rao, S. C., Kim, D.-S., Somers, D., & Sur, M. (1996). Subthreshold facilitation and suppression in primary visual cortex revealed by intrinsic signal imaging. *Proceedings of the National Academy of Sciences of the United States of America*, 93(18), 9869–9874.
- Tzvetanov, T., & Dresp, B. (2002). Short-and long-range effects in line contrast integration. *Vision Research*, 42(22), 2493–2498.
- Ullman, S., Gregory, R., & Atkinson, J. (1992). Low-Level Aspects of Segmentation and Recognition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 337(1281), 371–378.
- Walker, G. A., Izumi, O., & Freeman, R. D. (2000). Suppression outside the classical cortical receptive field. *Visual Neuroscience*, 17(3), 369–379.
- Walker, G. A., Ohzawa, I., & Freeman, R. D. (1999). Asymmetric suppression outside the classical receptive field of the visual cortex. *Journal of Neuroscience*, 19(23), 10536–10553.
- Watt, R., Ledgeway, T., & Dakin, S. C. (2008). Families of models for Gabor paths demonstrate the importance of spatial adjacency. *Journal of Vision*, 8(7), 1–19.
- Wertheimer, M. (1912). Experimentelle Studien über das Sehen von Bewegung. *Zeitschrift für Psychologie*, 61, 161–265.
- Wu, C.-C., & Chen, C.-C. (2010). Distinguishing lateral interaction from uncertainty reduction in collinear flanker effect on contrast discrimination. *Journal of Vision*, 10(3), 1–14.
- Yantis, S., & Nakama, T. (1998). Visual interactions in the path of apparent motion. *Nature Neuroscience*, 1(6), 508–512.
- Yen, S.-C., & Finkel, L. H. (1998). Extraction of perceptually salient contours by striate cortical networks. *Vision Research*, 38(5), 719–741.
- Zelinsky, G. J. (2008). A theory of eye movements during target acquisition. *Psychological Review*, 115(4), 787–835.
- Zenger, B., & Sagi, D. (1996). Isolating excitatory and inhibitory nonlinear spatial interactions involved in contrast detection. *Vision Research*, 36(16), 2497–2513.
- Zhaoping, L., & May, K. A. (2007). Psychophysical tests of the hypothesis of a bottom-up saliency map in primary visual cortex. *PLoS Computational Biology*, 3(4), e62.
- Zhaoping, L., & Snowden, R. J. (2006). A theory of a saliency map in primary visual cortex (V1) tested by psychophysics of colour–orientation interference in texture segmentation. *Visual Cognition*, 14(4-8), 911–933.

Chapter 2

The role of eye movements in a contour detection task

This chapter has been published as:

Van Humbeeck, N., Schmitt, N., Hermens, F., Wagemans, J., & Ernst, U. A. (2013). The role of eye movements in a contour detection task. *Journal of Vision*, 13(14), 1-19.

2.1 Introduction

Images that enter our eyes are first processed by neurons sensitive to only a small region of visual space and certain features (e.g., orientation, color). Crucial to our understanding of visual perception is the process by which the visual system groups these local neural responses to signal the presence of spatially coherent structures (for a broad review, see Wagemans et al., 2012). A powerful tool to achieve a better understanding of visual grouping is the contour detection paradigm of Field, Hayes, and Hess (1993). In this paradigm, a new type of stimulus was proposed to investigate the grouping of individual elements into global contours. In a typical contour detection experiment, observers are asked to detect the presence of a contour composed of spatially separate Gabor elements embedded in a background of similar but randomly oriented elements. The spatial separation of the distractor elements matches the separation of the contour elements, and therefore the contour can only emerge from the background due to the collinearly arranged contour elements forming a smooth path (for a recent review, see Hess, May, & Dumoulin, 2013).

The ability of human observers to detect a contour appears best when the contour forms a straight line and decreases as the degree of curvature of the contour, defined by the angle between the successive elements, increases (Field et al., 1993; Watt, Ledgeway, & Dakin, 2008). To account for this finding, Field et al. (1993) proposed that oriented elements interact with neighboring elements through a local “association field”. Typically, an association field indicates how strongly an element in the image can be collinearly grouped with its neighboring elements by assigning an association strength to each element in the field (May & Hess, 2007; Watt et al., 2008). The association strength of an element is high when the element is collinearly aligned with nearby elements and low when the distance, curvature, and misalignment from co-circularity between the element and its neighboring elements increases. This concept has been implemented in neural interaction models, involving facilitatory connections between nearly collinear elements and inhibitory interactions between parallel and orthogonal elements (Li, 1998; Mundhenk & Itti, 2002; Yen & Finkel, 1998). Recently, Ernst et al. (2012) developed an association field model which

captures human contour detection behavior to a previously unprecedented degree. This model has a well-defined probabilistic interpretation (Williams & Thornber, 2001) and uses an approach that reduces contour integration to an optimal inference problem. More specifically, it computes for each edge element in a stimulus the probability that this edge is part of a contour with certain statistical properties, thereby generating a probability map for contours of increasing length as processing time proceeds.

Studies making use of the contour detection paradigm typically limit the presentation time of the contour stimulus to reduce the influence of eye movements during the task. However, these short presentation times come at a cost. For example, Field et al. (1993) found that observers' detection performance became worse when the presentation time of the contour stimulus was reduced from 1 s to 0.25 s. The absence of eye movements may play an important role in this reduction. Given that the contour path did not always fall in the foveal region of the stimuli, preventing observers to make eye movements could have severely constrained contour detection. The role of eye movements in contour detection is further supported by findings showing that contour detection performance is highly dependent on the peripheral position of the contour (Hess & Dakin, 1997; May & Hess, 2007; Nugent, Keswani, Woods, & Peli, 2003). To account for such effects, May and Hess (2007) proposed a type of association field model in which the association fields are small in the foveal region of the stimulus and large in the periphery. Such a model predicts that elements of highly curved contours located far from the fixation point are more likely to be linked with distractor elements, due to larger association fields in the periphery linking elements over larger distances. Consequently, contours located in the periphery are less detectable, especially when they are curved. Eye movements can considerably support this contour integration process, allowing observers to foveate image elements that may be part of a contour. More specifically, we hypothesized that observers fixate regions of the image that are characterized by high association strengths. Under this assumption, measurements of fixation locations should reveal the internal association field used by observers during contour integration, providing a powerful tool to test models of contour integration which predict a specific association field. We here describe two experiments aimed to test this hypothesis by applying the association field model of Ernst et al.

(2012) to examine whether the model's predictions of salient collinear locations in an image match the fixation locations in a contour detection task.

Information processing during contour integration may also be reflected in other spatial and temporal aspects of eye movements. Fixation duration and saccade amplitude (i.e., the distance covered by a saccadic eye movement) have been proposed to indicate distinct modes of global and local processing during image viewing (Pannasch, Helmert, Roth, Herbold, & Walter, 2008; Unema, Pannasch, Joos, & Velichkovsky, 2005; Velichkovsky, Joos, Helmert, & Pannasch, 2005). For instance, in a visual search study, Over, Hooge, Vlaskamp, and Erkelens (2007) observed an increase in fixation duration and a decrease in saccade amplitude with viewing time. They hypothesized that this time course is the result of a search strategy of the visual system to gradually move from a global search to a more local search at finer spatial scales. The strategy did not depend on whether the exact appearance of the target and background was known in advance or not, suggesting that it reflected a general oculomotor strategy. However, a scene perception study of Mills, Hollingworth, Van der Stigchel, Hoffman, and Dodd (2011) showed that certain task requirements influence these oculomotor parameters differently, indicating that they are controlled by independent mechanisms instead of one intrinsic coarse-to-fine process (Pannasch et al., 2008; Unema et al., 2005). In this study, it was found that fixation duration gradually increased toward an optimal level of local processing, while saccade amplitude remained relatively stable over time. According to the authors, the size of saccades during a trial depended mainly on whether it was needed to acquire visual information across the entire scene. For instance, saccade amplitude remained high over the course of a trial in a search task in which the target was extremely difficult to detect. One possibility is that the time courses of fixation duration and saccade amplitude, instead of reflecting a built-in coarse-to-fine mechanism, can be strategically adjusted to the difficulty of the task. For instance, a study of Vlaskamp, Over, and Hooge (2005) suggests that fixation duration and saccade amplitude are closely related to the difficulty of finding a target during visual search. These authors found that fixation duration increased and saccade amplitude decreased with increasing target-distractor similarity (i.e., decreased target saliency), which reflects a more local processing strategy.

In sum, research on contour integration suggests that eye movements can substantially support contour detection, presumably by foveating likely contour candidates in the image. Moreover, certain saccadic characteristics such as fixation duration and saccade amplitude seem to be influenced, to a certain degree, by the saliency of a target during search, and by the time course of image processing. The purpose of this study was to examine the involvement of eye movements in the process of contour integration, especially in situations in which contours are difficult to find. We investigated how durations of individual fixations and saccade amplitudes vary as a function of time and difficulty of the contour integration task. In addition, we tested whether an association field model, proposed recently to explain contour integration (see Ernst et al., 2012), can predict fixation locations. Our hypothesis was that subsequent fixations are preferentially directed towards “hotspots” of association strength predicted by an association field. Participants were asked to find a contour hidden in a dense arrangement of oriented Gabor patches and to perform two tasks, while their eye movements were tracked. These included indicating whether the contour, which was always presented, was on the left or right of the visual field, or to determine whether a contour was present in the Gabor field. Task difficulty was manipulated by varying the path angle of the contour, which influenced the saliency of the contour elements.

2.2 Methods

2.2.1 Participants

Twenty-one observers participated in each task of the experiment, with normal or corrected-to-normal vision (age range 17-35). Two were authors, while the others were psychology students at the University of Leuven. All gave their written informed consent before participating in the experiment, which was approved by the local ethics committee.

2.2.2 Apparatus

The stimuli were displayed on a 22 inch CRT monitor (Iiyama HM204DT A) at a refresh rate of 75 Hz. A Pentium PC (NVIDIA GeForce 7600 GT graphics card) controlled the presentation of the stimuli, while a second PC recorded the eye movement data. Eye movements were measured using the Eyelink II video-based eye tracker (SR Research, Osgoode, ON, Canada), which uses two small cameras mounted to a head-band worn by the participants. The Eyelink II has a reported average accuracy of less than 0.5° , which is below the average distance of the elements in our display (i.e., 0.7°), allowing the accurate assignment of fixations to the individual elements in the display. To ensure that participants remained at a fixed distance from the screen and to avoid head movements which could possibly result in drift in the recorded eye positions due to head band slippage, a chin rest was used, positioned at 60 cm from the CRT screen. The eye tracker recorded the horizontal and vertical eye positions for both eyes at a rate of 250 Hz. Where possible, the combined pupil plus corneal reflection mode was used. For three participants, this mode resulted in clear distortions in the recorded eye position, because of reflections of the IR illumination to glasses or contact lenses. For these three participants, the pupil only mode was used instead. Participants' responses were registered by means of a computer keyboard.

2.2.3 Stimuli

The target stimulus (590 by 590 pixels; 23°) consisted of Gabor patches that were in cosine phase, with wavelength λ (0.20°), Gaussian envelope of SD $\lambda/2$ (0.10°), and Michelson contrast of 50% (see also Watt et al., 2008). An example of a stimulus can be found in Figure 2.1, which shows the stimulus sequence during a trial. Each stimulus with a contour target contained a single embedded contour placed in a background of randomly oriented Gabor elements. In the left-right decision task (in short: LR-task), Gabor paths were required to be at a distance of at least 50 pixels (2°) from the center of the screen, so that they were clearly located either in the left or right half of the Gabor field. The contours in the contour-present conditions of the present-absent task (in short: PA-task)

could appear in any position of the display. Each contour was defined by a sequence of seven Gabor patches whose orientations were set to an underlying contour “spline” (sequences often referred to as “snakes”; see inset of Figure 2.1). The positions of each of the Gabors in the path were chosen such that the mean distance between adjacent elements in the contour was the same as the mean distance between adjacent elements in the background. In order to manipulate the saliency of the contour and therefore task difficulty, the path angle of the contour was systematically varied (May & Hess, 2008; Watt et al., 2008). The path angle is defined as the angle between successive elements in the contour. For example, in the case of three elements, the path angle is defined by the difference in orientation between the lines formed by linking element 1 to element 2, and element 2 to element 3. Four different levels of path angle (0° , 5° , 10° or 20°) were used. The sign of the path angle could vary randomly within the contour. However, the orientation of the Gabor in a path was always aligned with the path angle (i.e., no additional jitter was used). As indicated earlier, the remainder of the Gabor field was filled with randomly oriented elements placed at random positions, in such a way that the mean distance between the center of each element and the center of any other neighboring element was 0.7° . Stimuli in the contour-absent condition of the PA-task did not contain a salient contour and consisted only of randomly placed Gabor elements.

2.2.4 Procedure

The experiment started by instructing the participants about the two tasks and by determining their dominant eye by asking them to look through a cylindrical object. In the first task, observers were instructed to indicate whether the contour was embedded left or right in the stimulus field (LR-task). In the second task, they were asked to indicate whether the contour was present or not (PA-task). In both tasks, either response was required in equal proportions. Participants responded by pressing either the letter “q” (“left” or “absent”) or the letter “p” (“right” or “present”) on a standard QWERTY keyboard. Before the experiment and on regular times during the experiment (after each block of 40 trials, if required), a calibration procedure for the eye tracker was performed. Calibration was repeated until all recorded fixations were aligned on a three by three grid

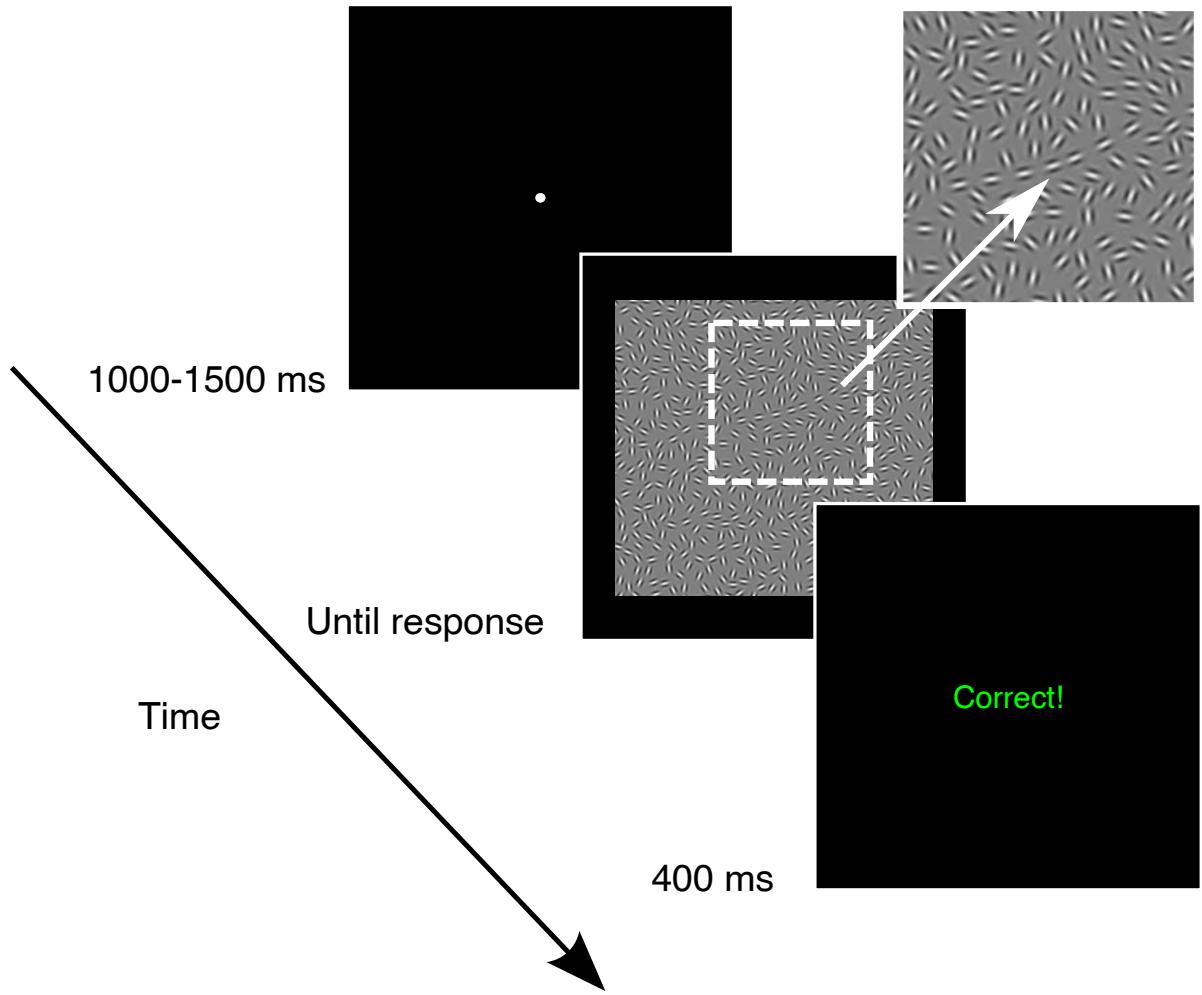


Figure 2.1. Illustration of the stimulus sequence. A fixation symbol was presented for a randomly selected duration between 1000 and 1500 ms. The target array was then presented until participants pressed a key to indicate whether a salient contour was located in the left or right half of the stimulus field (LR-task) or whether it was present or absent (PA-task). Next, feedback was provided for 400 ms. The inset shows a close-up of the contour path consisting of seven collinear elements.

for both eyes and the recordings of the first and the last fixation were at close proximity. Each trial started with a fixation dot, which was presented until the experimenter pressed the space bar to correct for drifts in the eye movement recordings due to small movements of the head. A second smaller fixation dot followed the drift correction, which was presented for a random duration between 1000 and 1500 ms. Participants were then presented with the stimulus image until they pressed one of two keys on the computer keyboard to indicate their response. During stimulus presentation, they were free to shift their gaze across the display. After each trial, feedback was provided for 400 ms indicating

whether the participant's response was correct or not (see Figure 2.1 for an example of the stimulus sequence during a trial). After each 40 trials, participants received a message indicating how many trials were remaining. For the LR-task, 60 different Gabor fields for each of the four path angles were used, resulting in 240 stimuli per participant. For the PA-task, 80 random Gabor fields were used in which no salient contour was present and 20 stimulus fields were used for each of the four path angles, resulting in 160 stimuli per participant. All the participants received the same Gabor fields, so that eye traces for each of the stimuli could be directly compared across the participants. The order of the trials was randomized for each participant.

2.2.5 Eye movement analysis

The default settings of the Eyelink II eye tracker were used to detect saccades, based on a velocity threshold of $30^\circ/\text{s}$ and an acceleration threshold of $8000^\circ/\text{s}^2$. While eye movement recordings were obtained for both eyes, we chose to focus on the recordings for the dominant eye (as determined with the cylindrical object viewing task) in the analysis. Fixations shorter than 50 ms and longer than 2000 ms, as well as saccades with amplitudes smaller than 0.1° , were excluded from data analysis. The saccade before and the saccade after an excluded fixation were merged to form a new saccade and the same method was used for fixations before and after removed saccades. The first fixation in each trial was discarded, since it results from the preceding fixation dot at the centre of the screen used for drift correction. Across all trials of this study, average fixation duration was 275 ms ($SD = 133$) and average saccade amplitude was 5.41° ($SD = 4.56$).

Statistical analyses were performed on the individual eye movements, which were in total 37,352 in the LR-task and 32,193 in the PA-task. Eye movements were nested within images and within observers for both tasks. The images were crossed with individuals, since each observer saw the same set of Gabor field stimuli. Given that our dataset has a hierarchical structure, multilevel models were used in which images and observers were treated as crossed random factors (Hoffman & Rovine, 2007; Locker, Hoffman, & Bovaird, 2007; see Mills et al., 2011 for a similar analysis on eye movement data). An important

advantage of multilevel analysis is that it allows for an unbalanced design, as is the case for our dataset, in which there are differences in the timing and number of eye movements across observers and trials. Models were estimated using maximum likelihood via the MIXED procedure of SPSS (syntax used to estimate the models is available from the authors upon request). Although the methods are relatively robust against violations of the normality assumption of the underlying distribution, we applied base 10 logarithmic transformations to our data to obtain normal data distributions.

To examine how fixation duration and saccade amplitude changed over the course of a trial, we tested a quadratic model of change over time for these two parameters, as exploratory analyses of our data showed a quadratic change over time in fixation duration and saccade amplitude. The time for each trial was first normalized to the interval starting at the time when the first saccade was initiated in the center of the screen, and ending when the last saccade ended. This interval was divided into 10 subintervals of equal width, and fixations sorted into these subintervals according to their normalized starting time. Statistical tests were performed separately for the LR-task and the PA-task. We included two between-trial predictors. The first predictor was the saliency of a contour. Conditions included the different path angle conditions in both tasks and the contour-absent condition in the PA-task. The second predictor is the correctness of the given manual response. In addition, we tested whether these predictors interacted with the linear and quadratic effect of time.

2.2.6 Architecture of contour integration model

We compared our experiments with predictions from the association field model by Ernst et al. (2012). This contour integration model was realized by a two-dimensional layer of $i = 1, \dots, N$ neuronal populations which were recurrently connected by synaptic weights w_{ij} . For reducing computational complexity, only populations i were used whose receptive fields were centered on the edge elements in a particular stimulus, and which had the same preferred orientations.

In order to integrate neighboring edges, the synaptic weights were chosen such that a

stronger interaction between nearby element pairs i and j results when these are arranged in a collinear and co-circular fashion, thus realizing an association field. The weights were sampled from a function $A(r, \alpha, \beta)$ defined as

$$A(r, \alpha, \beta) \propto \exp(-r/\lambda) \cosh(1/\sigma_\alpha^2 \cos(\beta/2 - \alpha) + 4/\sigma_\beta^2 \cos(\beta/2)) . \quad (2.1)$$

The radial part of this function decays exponentially with the length constant λ depending on the distance r between two edge elements. The angular part depends on the angles α and β , with β referring to the angle between the orientations of two edge elements, and α denoting the angle of the destination edge i from the collinear continuation of the origin edge j . σ_α and σ_β are the scaling constants that determine the form of the association field (panel A of Figure 2.2). By setting both parameters to small values, contours with quite straight curvatures are favored during integration. By setting the parameters to higher values, contours with larger curvature can also be integrated, but with the penalty of making the model less robust against noise. Panel B of Figure 2.2 shows a typical example of an association field with the parameters chosen for our simulations. Note that the association field is not symmetric and extends into a particular direction. Unidirectionality of the association field turned out to be beneficial for contour integration (Ernst et al., 2012), but it requires that each edge element in a stimulus is represented by two populations i and i' with association fields extending into opposite directions.

To account for reduced saliences of edge elements i in peripheral vision (Foley, Varadharajan, Koh, & Farias, 2007), an additional scaling factor b_i for the afferent input to each population i was introduced. In the contour detection study of Ernst et al. (2012), it was found that b has to decrease with distance E from the fixation spot for the model to be able to successfully explain human contour integration. The required functional dependency of b on eccentricity E was found to be compatible to earlier psychophysical studies (e.g., Foley et al., 2007), and is quantified by

$$b(E, \mu, \nu) = 1 + 2\mu \left(\left(\frac{E}{E_{max}} \right)^\nu - \frac{1}{2} \right) . \quad (2.2)$$

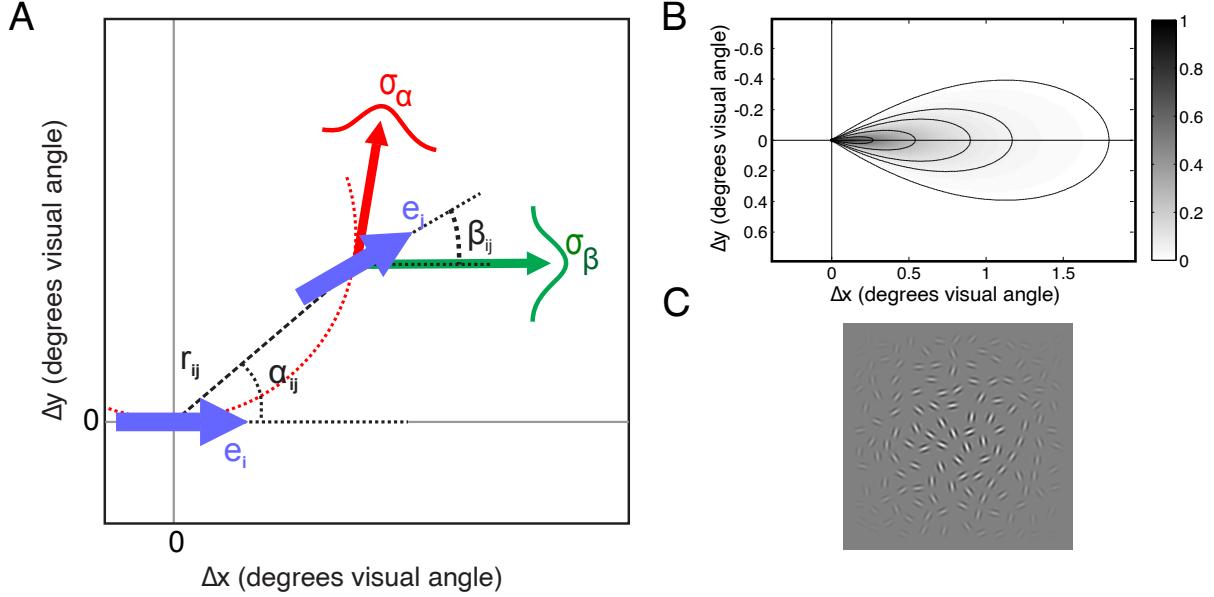


Figure 2.2. Parameters and shape of the association field and eccentricity scaling. (A) The geometrical relationship between edge elements e_i and e_j is defined by the distance r_{ij} between the elements and the angles α_{ij} and β_{ij} . α_{ij} refers to the angle of edge e_j from the collinear continuation of edge e_i and β_{ij} is the angle between the orientations of the two edges. The red arrow specifies the direction e_j should have for perfect co-circularity between e_i and e_j . The association field decreases on a length scale σ_α with increasing deviation of the orientation of e_j from this direction. The green arrow depicts the case in which the two edge elements would have the same orientation. The association field decreases on a length scale σ_β with increasing deviation of the orientation of e_j from this direction. Figure adapted from Ernst et al. (2012). (B) The association field, defined as the product of an angular part and a radial part (see Methods section), for an element positioned at the origin and with an orientation equal to zero. It is defined by parameter values fitted to reproduce human contour detection behavior in our LR-task. Grey scale values specify the association strength, with darker shades indicating locations that give rise to higher association strengths. (C) The association field model also contains a scaling factor b_i for each edge e_i which determines its visibility. In the image, the degraded visibility of the contour and other elements located in the periphery would be reflected in the model by relatively low b_i values for these elements.

Here μ describes how strongly b varies with eccentricity and ν regulates how steeply b varies with eccentricity. For $\nu < 1$, b is concave down and for $\nu > 1$, b is concave up. Panel C of Figure 2.2 illustrates an example for such eccentricity scaling by modulating the contrast of background and contour elements located in the periphery. Although it may be argued that no scaling for peripheral locations may be required in the presence of eye movements, we included the b parameter in the original fit of our data, specifically to test whether observers' eye movements were indeed sufficient to make collinear configurations in the periphery more salient.

2.2.7 Model dynamics

The dynamics of the contour integration model are described by a time-continuous differential equation (DEQ) for the activation $p_i(t)$ of population i , which is a generalisation of the time-discrete model used in Ernst et al. (2012):

$$\tau \frac{dp_i(t)}{dt} = -p_i(t) + \eta(t)g \left[b_i \sum_j w_{ij} p_j(t) \right]. \quad (2.3)$$

In this equation, $g[\dots]$ denotes a rectifying neural gain function, and $\eta(t)$ is a normalization factor which is defined by the sum over the synaptic input term in rectangular brackets,

$$\eta(t) = \sum_i b_i \sum_j w_{ij} p_j(t). \quad (2.4)$$

If contour integration is regarded as a probabilistic inference problem (Ernst et al., 2012), p_i can be interpreted as the likelihood that a contour passes through an edge element in a stimulus. p_i is computed iteratively for contours of increasing length. For retaining the analogy to neural networks and neural processes in visual cortex, however, we will refer to this variable as the model’s “activation” throughout the following text.

2.2.8 Fitting procedure

To avoid overfitting, we use one part of the available experimental data to perform the fit, and test the model’s predictive power on a different part of the data. In particular, we use the contour detection behavior of all subjects in the LR-task to optimize the parameters of the contour integration model, and then test the generalizability of the data fit by applying a model with the same parameters for subsequent prediction of saccades in the PA-task.

For our fit, we require the model both to reach or surpass human contour detection performances and to reproduce excess correlations among human observers. The term excess correlations refers to human decisions which are either less frequent or more frequent

than expected from human mean performance. One example shall illustrate this idea: assuming that performance of human observers in detecting a contour is 70% correct, we expect on average 14 correct decisions for each stimulus when we would have 20 subjects in total. If we find a stimulus with only one correct and 19 incorrect decisions, meaning that observers are systematically wrong, we also expect our model to “fail” on this particular contour. These two criteria were quantified in two measures Z and C , respectively. Z takes a value between [0, 1] and denotes the percentage of stimulus conditions in which a model reaches or surpasses mean human performance. We required our model to achieve $Z = 1$. C can also take a value between [0, 1] and is related to the integral of a ROC curve. It quantifies excess correlations between the correct decisions of two (sets of) observers: if $C = 0.5$, identical decisions of the two observers are fully explained by their mean performances. If the number of identical decisions of the two observers increases over this chance level, C becomes larger than 0.5 with a maximum attainable value of 1 (for details with regard to the precise computation of C , we refer the reader to Ernst et al., 2012). In order to quantify how well a model reproduces human behavior, we compare average excess correlations $C_H = 0.79$ among subjects with average excess correlations C_M between model decisions and subjects, requiring the model to come as close as possible to the “benchmark” value C_H .

Fitting was performed on five parameters, namely σ_α and σ_β (shape of association field), μ and ν (scaling of edge saliency with visual field eccentricity), and λ (scaling of coupling strength with edge distance). For each parameter, a plausible range of values was selected and divided into equidistant intervals. After initialization with a uniform activation distribution, the model was simulated for a time interval of $T = 10\tau$ for all possible parameter combinations. The model with $Z = 1$ and the highest value of C_M identified by this procedure was subsequently used to predict saccade targets.

2.2.9 Prediction of fixation locations

The model generates time series of activities $p_i(t)$ for all populations i . Predictions for saccade targets were derived by comparing p with a time-varying threshold $\Theta(t) =$

$\langle p_i(t) \rangle_i + 0.5\sigma(p_i(t))$, where $\langle \cdots \rangle$ denotes the mean and σ the standard deviation of the model activation *at a particular point in time*. If $p_i(t) > \Theta(t)$, the corresponding edge element i was considered a potential saccade target. For all stimuli containing a contour, we also removed all fixations within a radius $R_T = 3.5^\circ$ around the center-of-mass of the target contour. Removing fixations near the target had two reasons. First, we wanted to find out if the model can predict fixations *prior* to detection of the target contour. Second, we indirectly used this information already during parameter fitting because saccade trajectories end near a contour if it is detected by an observer.

For assessing how well the model predicts the remaining fixations, we count the relative number of fixation spots $t_{pos}(R)$ (“true positives”) which are within a range R of at least one potential saccade target identified by the model. This procedure is also performed for the same model prediction compared with fixations from a *different* stimulus, thus giving a relative number of “false positives” $f_{pos}(R)$. By relating t_{pos} with f_{pos} via the free parameter R , we obtain a receiver-operator characteristics (ROC) whose integral indicates how well the model can predict the fixation spots (0.5 is chance level).

The ROCs were first computed for each observer individually, and then averaged by taking the maximum value achieved over simulation time t for each observer. This procedure accounts for the possibility that different subjects might choose saccade targets from different stages of the contour integration process: while one observer makes saccades only to contours of at least five elements, a different observer might judge contours of already three aligned edge elements as sufficiently “interesting” for performing a saccade and for conducting a further inspection of its neighborhood.

2.3 Results

2.3.1 Task performance

Mean performance on the contour integration task decreased as the path angle of the target contour increased. For increasing path angles 0° , 5° , 10° and 20° , the average

proportion correct across observers equalled respectively 97%, 96%, 90%, and 74% in the LR-task and 92%, 88%, 77%, and 56% in the PA-task. For the condition in which no contour was present in the display, mean performance was 81% correct. Box plots of the number of fixations and response times as a function of path angle for both the LR-task and the PA-task are shown in the left and right panel of Figure 2.3, respectively. Repeated measures ANOVA revealed a significant difference between path angle conditions in log response times (Greenhouse-Geisser correction, $F(1.74, 34.86) = 224.18, p < 0.001$ for the LR-task and $F(1.71, 34.26) = 89.27, p < 0.001$ for the PA-task) as well as log number of fixations (Greenhouse-Geisser correction, $F(2.33, 46.65) = 170.58, p < 0.001$ for the LR-task and $F(1.79, 35.82) = 74.47, p < 0.001$ for the PA-task). Pairwise comparisons between path angle conditions showed that all differences were significant after Bonferroni adjustment (requiring $p < 0.008$ for the LR-task and $p < 0.005$ for the PA-task), except the difference in log number of fixations between the smallest path angles in the LR-task ($p = 0.074$).

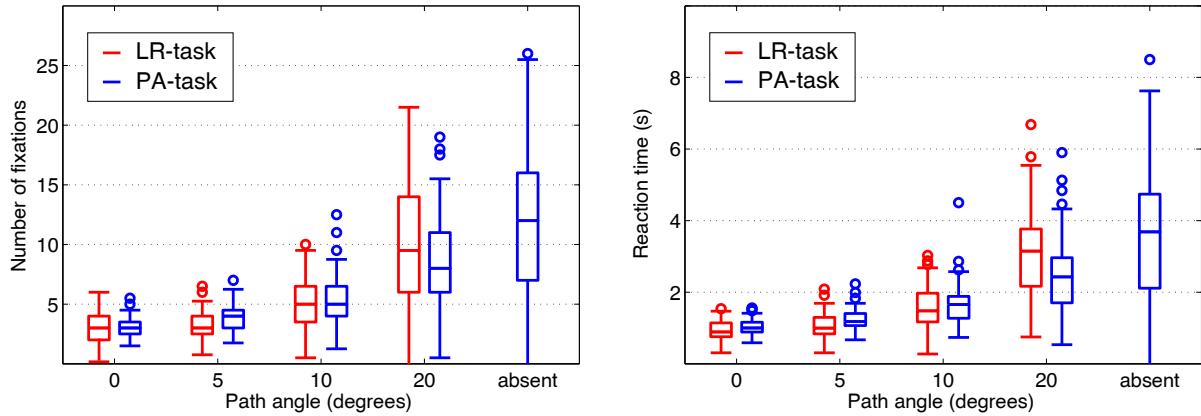


Figure 2.3. Box plots of number of fixations (left panel) and reaction times (right panel) as a function of path angle of the contour (including the condition in which no salient contour is present in the PA-task). As the path angle increases, number of fixations and reaction times increase. Data from the LR-task are shown in red, while data from the PA-task are shown in blue.

2.3.2 Eye movement behavior

To account for random variations between trials (i.e., images) and between observers, the regression coefficients in our multilevel models predicting fixation duration and saccade

amplitude were treated as random variables. For the prediction of fixation duration, significant sources of random variation after controlling for the predictors contour saliency and correctness were a random intercept and slope for observers and a random intercept for trials. For the prediction of saccade amplitude, we included a random intercept, slope and quadratic term for observers and a random intercept for trials.

The left and right panels of Figure 2.4 show the change in log fixation duration over the course of a trial as a function of contour saliency for the LR-task and PA-task, respectively. Fixation durations increased gradually, with the rate of change decelerating over time, and slightly decreased towards the end of the trial. The last fixation in each trial appeared considerably longer than the preceding fixations due to the overlap with the observers' responses and these fixations were excluded from further analysis of fixation duration. For both tasks, multilevel analysis showed a significant positive linear time effect ($F(1, 191.55) = 200.27, p < 0.001$ for the LR-task; $F(1, 832.71) = 230.82, p < 0.001$ for the PA-task) and negative quadratic time effect ($F(1, 32,322.09) = 183.82, p < 0.001$ for the LR-task; $F(1, 28,586.34) = 150.81, p < 0.001$ for the PA-task), indicating that the linear effect of time decelerated significantly over the course of the trial.

In addition, there was a main effect of contour saliency for the LR-task ($F(3, 2,324.74) = 20.20, p < 0.001$) and the PA-task ($F(4, 6,614.24) = 8.27, p < 0.001$). The effect size of contour saliency in predicting fixation duration was assessed using pseudo- R^2 statistics (Singer & Willett, 2003), indicating the change in the variance components after including the effect of contour saliency in the model. As to be expected, including the predictor contour saliency caused only a small change in the random variance of the intercept and linear time effect for *observers*, as each observer received all path angle conditions. However, contour saliency accounted for 51% and 62% of the random intercept variance for *stimuli* in the LR-task and the PA-task, respectively. Moreover, contour saliency did not interact with the linear and quadratic effect of time, meaning that the effect of contour saliency remained the same over the course of a trial. There were no significant differences in log fixation duration between correct and incorrect trials neither for the LR-task ($F(1, 32,589.16) = 1.07, p = 0.301$) or the PA-task ($F(1, 28,428.23) = 0.35, p = 0.553$). The fitted curves in the left and right panels of Figure 2.4 show the pattern of

change in log fixation duration across time intervals as predicted by a quadratic model of time which includes a main effect of contour saliency. In a separate analysis, we combined the datasets of the LR-task and the PA-task and found no difference in log fixation duration between the two tasks ($F(1, 52.84) = 0.62, p = 0.434$).

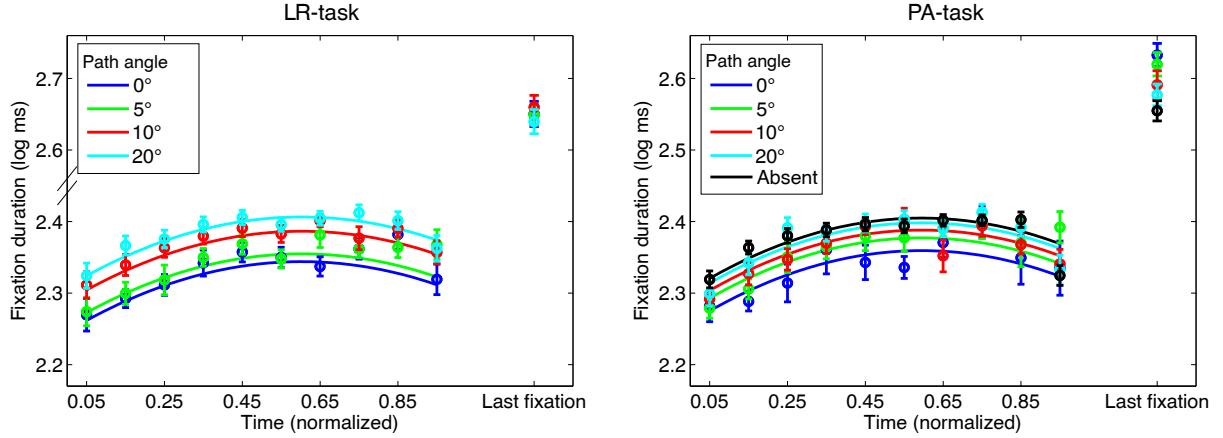


Figure 2.4. Log fixation duration in the LR-task (left panel) and the PA-task (right panel) as a function of normalized viewing time can be predicted by a quadratic model of time including a main effect of contour saliency, as indicated by the solid lines. The data fits suggest that fixation durations increased gradually, and slightly decreased towards the end of the trial. Different colors indicate the contour saliency levels (i.e., the different path angle conditions and the contour-absent condition of the PA-task, see legend). Error bars represent the standard error of the mean across observers.

Panel A of Figure 2.5 shows the effect of viewing time on log saccade amplitude as a function of contour saliency for the LR-task. The pattern of change in log saccade amplitude suggests a quadratic pattern of change, with saccade amplitude peaking at around 40% of the time in the trial and decreasing towards the end of the trial. Indeed, the dependence of saccade amplitude on time can be predicted by a quadratic model of viewing time, as shown by the fitted curve in panel A of Figure 2.5. Our analyses showed that the linear ($F(1, 24.12) = 17.13, p < 0.001$) as well as the quadratic time effect was significant ($F(1, 24.25) = 42.25, p < 0.001$). The main effect of contour saliency failed to reach significance ($F(3, 12,268.96) = 2.41, p = 0.065$).

Panel B of Figure 2.5 shows the dependence of log saccade amplitude on viewing time as a function of contour saliency and correctness of response for the PA-task. Here, log saccade amplitude seems to follow two trajectories: the blue curves depict the saccade amplitudes for trials in which no contour was perceived (i.e., trials in which observers

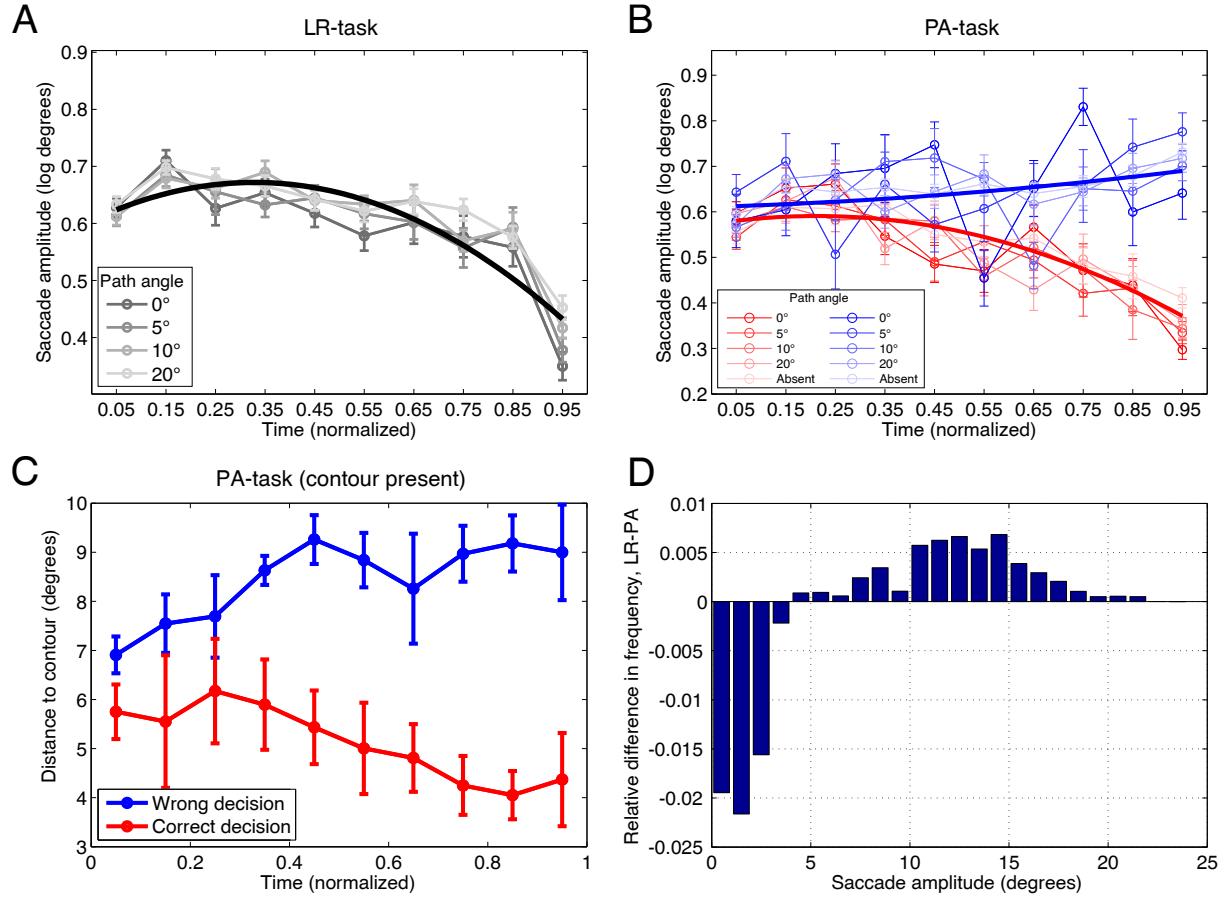


Figure 2.5. Log saccade amplitude in dependence on normalized time during a trial. Error bars indicate the standard error of the mean across observers. (A) For the LR-task, log saccade amplitude peaks at around 0.40 normalized time and decreases towards the end of the trial. This can be predicted by a quadratic model of time, as indicated by the bold black curve. Different shades of grey indicate different path angles of the contours. (B) For the PA-task, the time course of saccade amplitude depended on the percepts of the observers (i.e., whether they reported the contour to be present or absent). The blue curves show log saccade amplitudes (all path angles and contour-absent condition) when no contour was perceived, while the red curves show log saccade amplitudes when observers reported having perceived a contour. The time course of saccade amplitude can be predicted by a quadratic model of time in which the effect of perceiving the contour interacts with the linear and quadratic effect of time, as indicated by the bold curves. Different shades of blue and red indicate either different path angle conditions or the contour-absent condition. (C) Mean distance of fixation to the contour in dependence on normalized time during a trial for the contour-present conditions in the PA-task. Observers progressively made fixations closer to the contour in case of a correct decision, whereas they fixated further from the contour over the course of a trial in case of a wrong decision. (D) Difference in saccade amplitude distributions between the LR-task and the PA-task for trials in which a contour was present.

either correctly or incorrectly indicated the absence of a contour). In these trials, saccade amplitude gradually increased over the course of the trial. The red curves show saccade

amplitudes when observers reported finding a contour. Here, saccade amplitude followed roughly the same time course as in the LR-task. Thus, the time course of saccade amplitude seemed to depend on in which contour saliency condition (i.e., present or absent conditions) the observers gave a correct or false response. This is reflected by a nearly significant interaction between correctness, contour saliency and the linear time effect ($F(4, 32,058.75) = 2.35, p = 0.052$), and a highly significant interaction between correctness, contour saliency and the quadratic time effect ($F(4, 31,963.95) = 11.90, p < 0.001$). As a measure of effect size, including this interaction between contour saliency and correctness reduced the random intercept variance for stimuli by 57%. The bold red and blue curves in Panel B of Figure 2.5 show the fits of a quadratic model of time in which the effect of reporting a contour interacts with the linear and quadratic effect of time. When only considering trials in which a contour was present in the PA task, the distance of fixation to the contour follows a similar time course to that of saccade amplitude. Panel C of Figure 2.5 shows the mean distance to the contour as a function of normalized time. For trials in which they correctly perceived the contour, observers progressively made fixations closer to the contour, whereas they gradually fixated locations further from the contour for trials in which they did not perceive the contour. The datasets of the LR-task and the PA-task were combined into a further analysis and a significant difference in log saccade amplitude was found between the two tasks ($F(1, 52.06) = 8.01, p = 0.007$): saccade amplitude was on average larger for the LR-task than for the PA-task. The difference in saccade amplitude distributions between the LR-task and the PA-task is also obvious from panel D of Figure 2.5, where it can be seen that long saccades prevail in the LR-task.

2.3.3 Prediction of fixation locations

Before applying the association field model to our eye movement data, we performed a parameter fit of the model on the observed contour detection performance as described in the Methods section. An optimal fit was achieved when elements were assumed to have the same visibility over the whole display (flat eccentricity scaling function, $\mu = 0$). This result is consistent with the idea that elements and contours at the borders of the

screen, which would have lower visibility when observers were not allowed to make eye movements, become more salient when saccades are allowed to inspect locations in the periphery. For the association field parameters, which determine the coupling matrix w_{ij} , we found optimal values of $\sigma_\alpha = 0.15$, $\sigma_\beta = 0.24$, and $\lambda = 0.3$ (left panel of Figure 2.6).

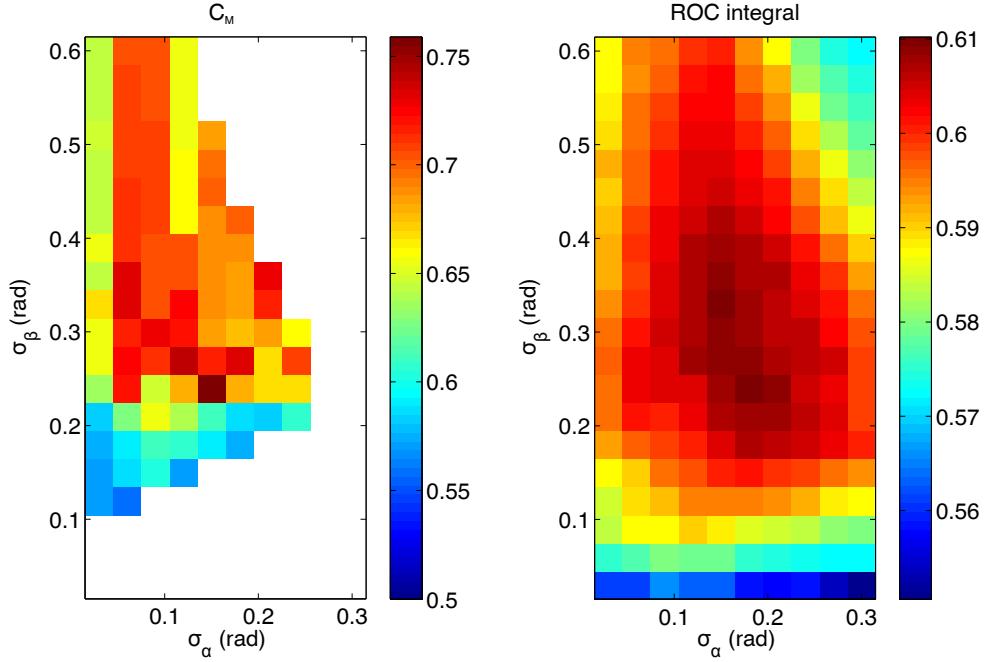


Figure 2.6. Fitting the model to psychophysical performance and eye movement data. Left panel: Excess correlations quantified by C_M for different models with association field parameters σ_α and σ_β varied independently. The other parameters were held constant at their optimal values. The color scale indicates how well the model reproduces human excess correlations. Parameter combinations for which model performance was inferior to human performance are left white. Right panel: ROC integral for different models with association field parameters σ_α and σ_β varied independently. The color scale indicates how well the model predicts human fixation locations.

To determine whether saccades are made to collinear structures that are salient according to our contour integration model, we compared actual fixations with model predictions. Figure 2.7 shows such a comparison for four examples from each of the experimental conditions. For the purpose of visualization, we constructed an activity distribution from the discrete activations p_i by using a convolution with a Gaussian profile (half-width 2 degrees of visual angle). The neural activity distribution from the association field is shown as a normalized heat map, with red indicating high activity. Fixations of all subjects are superimposed on these activity maps, shown as black stars. It can be seen

that hotspots of model activity in general match well with dense clusters of fixation spots, but there are also examples where participants fixate regions of low model activity. This impression of the data from these examples is confirmed by a correlation analysis for all stimuli in the PA-task where contours are absent (Figure 2.8). For this analysis, fixation locations and model predictions for each stimulus were first convolved with a Gaussian profile of half-width one degree of visual angle. Then, a correlation coefficient over the resulting two-dimensional distributions was computed. The distribution of correlation coefficients from this data set (blue bars in Figure 2.8) is centered around 0.3, and well separated from the distribution of coefficients for surrogate data, for which we compared model predictions to fixations made for a randomly selected, *different* stimulus (red bars in Figure 2.8).

As a more direct quantitative measure of the predictive power of the model, we computed a receiver-operator characteristic (ROC), which relates the amount of “true positives” to “false positives” in dependence on a decision criterion R . In this case, R is given by the radius around a candidate edge in a stimulus. A candidate edge is an edge for which the model predicts an activation which is higher than a certain threshold (see Methods). If a fixation spot falls within a radius R around a candidate edge, it is taken as a “true positive”. If a fixation spot of a saccade trajectory for a randomly chosen, different stimulus (with the same contour path angle) falls within a radius R around a candidate edge, it is taken as a “false positive”. Loosely speaking, the integral over the ROC then quantifies how well fixation spots are predicted by the model compared to “chance” prediction performance.

In this analysis, we obtained a value of 0.608 for the LR-task, a value of 0.64 for the PA-task when the contour was present and a value of 0.583 for the PA-task when the contour was absent. These numbers are significantly different from the value that would be obtained by chance (i.e., 0.5, with a threshold of 0.525 for $p < 0.05$). The ROC analysis therefore shows that the model’s performance is clearly above chance level, meaning that activation in an association field model predicts fixations in a contour-detection task well. However, not all human saccades can be explained on the basis of the model. Predictions are worse for stimuli containing only background elements. For these stimuli, the high

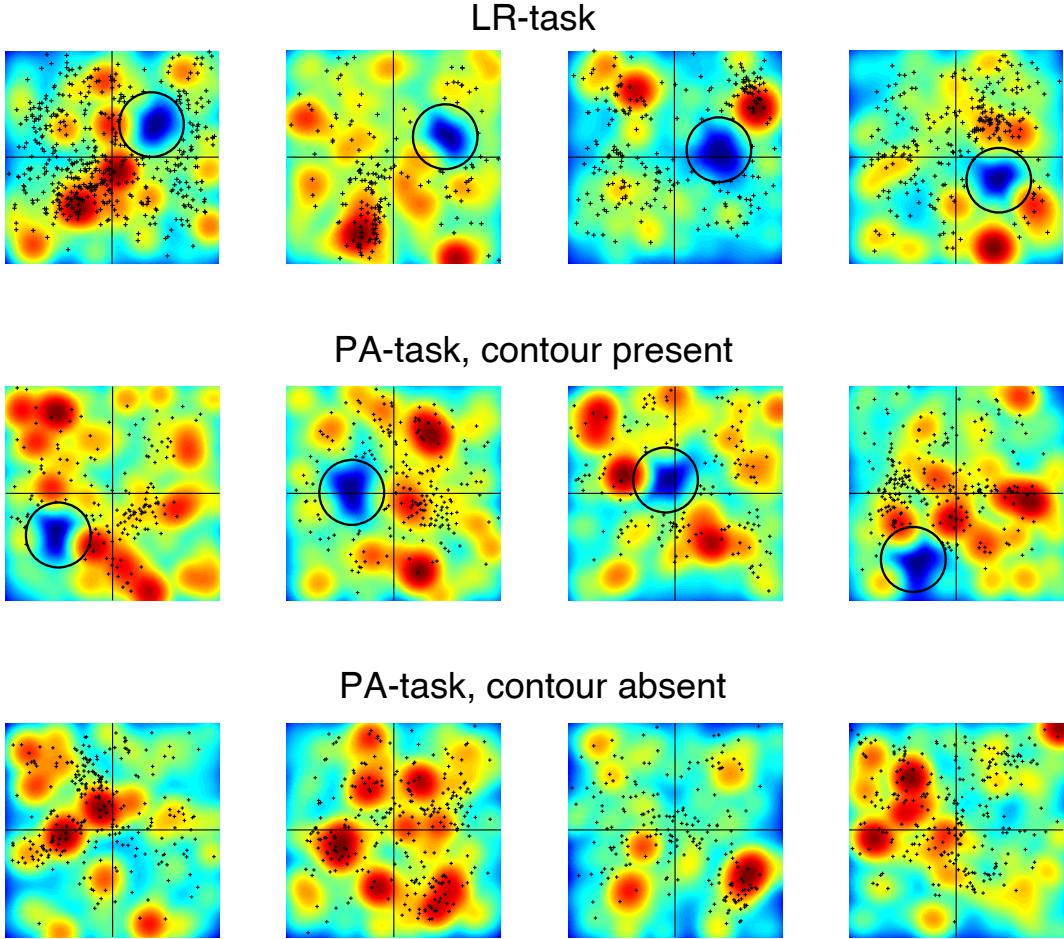


Figure 2.7. Heatmaps of model activity for a subset of the stimuli in the LR-task (top row), the contour-present condition in the PA-task (middle row), and the contour-absent condition in the PA-task (bottom row). Fixations of all observers are indicated as black stars. High activity is colored in red, while low activity is displayed in blue. The black lines indicate the vertical and horizontal meridian of the display. The black circle encloses the region around the target contour from which fixations were removed before comparing the fixation locations to model predictions. The plots suggest that fixation spots in general tend to cluster at locations of high model activity, although in some cases participants fixate regions where model activity is low.

percentage of false positives emerging in a confined stimulus arrangement with a large number of fixations is problematic, reducing the ROC integral considerably. Hence, the reported ROC values may underestimate the true predictive power of the model. As a consistency check for our initial model fit, we performed an independent search to identify parameter sets maximizing the ROC integral values. This procedure revealed a similar optimal parameter range as found by the fit using human performance and excess correlations as the relevant measures (right panel of Figure 2.6).

These analyses can be extended to individual participants, as illustrated in the top left

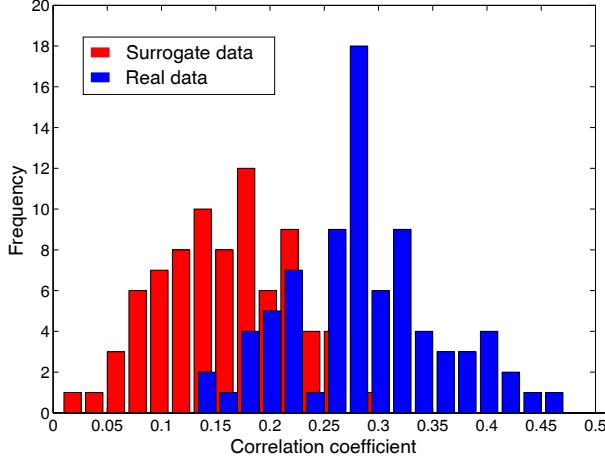


Figure 2.8. Distributions of correlation coefficients indicating how well our model predicts fixation locations for real (blue) and surrogate (red) data of the contour-absent condition in the PA-task. It can be seen that the distribution of correlation coefficients for the real data is well separated from the distribution of coefficients for the surrogate data.

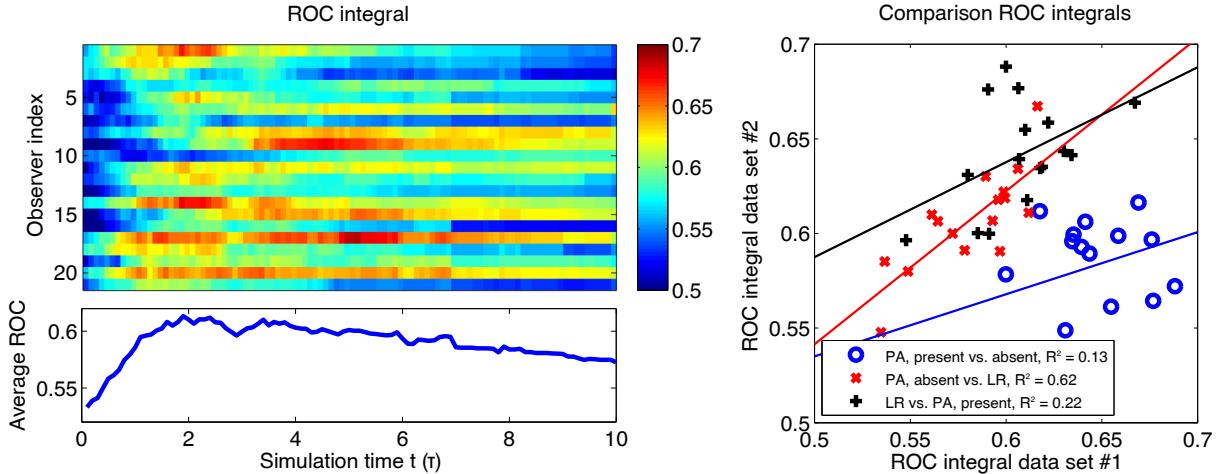


Figure 2.9. Left panel: ROC integral, indicating the performance of the optimal model in predicting eye fixations in the PA-task in the contour-absent condition, for different observers in dependence on simulation time t . Overall performance of the model and the time interval in which the model has the highest predictive power differ greatly between subjects. ROC integrals averaged across observers as a function of simulation time are shown in the lower graph. Right panel: Correlation between the ROC integrals of observers in the LR-task and the contour-present and contour-absent condition of the PA-task.

panel of Figure 2.9 showing ROC integrals for each of the individual observers over time. Whereas the model predicts fixations for certain subjects well, for others its predictions barely exceed chance level. This analysis also suggests that the maximum predictive performance for each observer is reached at different points in (simulation) time. This is reflected by the absolute values on the time axis, measured in units of the model's time

constant τ , which can be regarded as an indicator of how many elements the model has integrated into a potential contour at these points in time. For example, a time of 2τ means that any combination of three edge elements in the stimulus has been integrated by evaluating any set of two links between it. As observers' maxima range from about 2τ to 5τ , it seems that human saccades are best predicted by identifying contour configurations of 3 to 6 elements. This is also evident from the bottom left panel of Figure 2.9 which shows the average ROC integral across observers as a function of time. The right panel of Figure 2.9 demonstrates that the individual differences in predictive power were preserved across experiments. If fixations of an individual observer were well predicted in the LR-task, they were also well predicted in the PA-task (both for contour-present and contour-absent conditions). For the stimuli in which a contour was present, we also found a negative correlation between ROC integral and mean number of fixations per observer (slopes -0.0053 and -0.0056 for a linear regression with $R^2 = 0.36$ and $R^2 = 0.30$ for the LR-task and PA-task, respectively, data not shown).

2.4 Discussion

We conducted two contour detection tasks to investigate how saccadic eye movements and contour integration combine to support a perceptual decision. In both tasks, observers were free to move their eyes across the image to search for the contour. The difficulty of grouping contour elements into a global percept of a contour was manipulated by the degree of curvature (path angle) of the contour, while the number of elements in the contour was kept constant. Main results common to both tasks were:

- Gabor fields containing a relatively salient contour (i.e., with a small path angle) led to small numbers of eye movements which each had short fixation durations, resulting in fast response times. The opposite was found for Gabor fields containing a contour with a large curvature, where large numbers of fixations and longer fixation durations, as well as long response times were found.
- For all path angles, fixation durations increased during the course of a trial, whereas

the time course of saccade amplitudes depended on whether or not a potential contour could be perceived. When subjects indicated that they could not identify a contour (independently of whether a contour was present or absent), saccade amplitudes increased over a trial. When participants reported perceiving a contour, saccade amplitudes decreased.

- Many fixations could be predicted on the basis of saliency maps obtained from an association field model of contour integration (Ernst et al., 2012). This suggests that measurements of fixation locations can be used to probe the observers' internal representation of the association field. Fitted model parameters on the basis of behavioral responses indicated that saccades compensate for the decrease in contour saliency usually observed in peripheral vision.

These results will be discussed in more detail in the following sections.

2.4.1 Effect of contour saliency

The strongest effects of the saliency of the contour were found on the number of fixations and the overall reaction time, which both increased with the path angle of the contour. The highest reaction times and number of fixations were observed for the condition in which no salient contour was present in the PA-task. In addition, we found a significant increase in fixation durations with decreasing contour saliency in both the LR-task and the PA-task. As the observed variations in fixation duration are much smaller, the reaction time is mainly determined by, and thus proportional to, the number of fixations. This finding has also been observed in other studies on visual search (McCarley et al., 2006; Zelinsky & Sheinberg, 1995).

The effect of path angle on fixation durations is supported by previous visual search studies showing increased fixation durations with decreasing target saliency (Hooge & Erkelens, 1996; Jacobs, 1986). For instance, Hooge and Erkelens (1996) asked observers to search for a circle among six letter Cs (i.e., rings containing a gap) and found longer fixation durations when the distractor elements had smaller gap sizes, i.e., when the target was less salient. Notably, the increase of fixation duration for higher path angles in

our study existed from the beginning of a trial, despite the fact that trials with different contour path angles were randomized and observers could not know the saliency of the contour in a given trial in advance. As the background itself did not reveal any information about the path angle of the embedded contour, it is an indication that on-going contour integration processes had already extracted information about the contour (possibly computing an initial, but still incomplete saliency map), and that this information was used to directly control the temporal parameters of the eye movements. In consequence, more time is spent on examining edge configurations in the vicinity of a fixation location when no clear contour candidates are present in the initial assessment of a stimulus. Hence, fixation durations seem to be under the immediate control of the information available in the stimulus, consistent with evidence showing that the duration of some fixations increased when the onset of the stimulus at the saccadic landing position was delayed (Henderson & Pierce, 2008; Henderson & Smith, 2009). Contrary to our finding, Hooge and Erkelens (1996) observed that target saliency did not influence fixation durations when it changed from trial to trial. According to the authors, the unpredictability of task difficulty prevented observers to correctly adjust their fixation durations because estimation of time needed at fixation was based on previous trials. A possible reason for the difference between our findings and those of Hooge and Erkelens (1996) may be the amount of processing required to solve the task. Our tasks were relatively complex and involved the grouping of local elements into contours, whereas the study of Hooge and Erkelens (1996) involved a much simpler search task in which the target only differed from background elements with respect to one feature (i.e., a gap). In our study, it might be more important to adjust fixation durations at the beginning of each trial so that more processing resources can be allocated to fixation locations in difficult trials. In fact, Hooge and Erkelens (1996) acknowledged the potential importance of the first fixation as the time when the difficulty of the current trial is assessed. It should be noted that fixation durations were approximately the same for correct and incorrect trials, indicating that fixation duration did not depend on whether a correct perceptual decision was made at the end of the trial. Moreover, we found no difference in fixation durations between the LR-task and the PA-task. This is not surprising, since the contour stimuli used in the two tasks were very similar, requiring the same amount of processing at fixation location.

Our stimuli did differ between tasks in spatial constraints related to the location of the contour, which influenced the amplitude of saccades (see below).

2.4.2 Dynamics of eye movements

Fixation durations increased in the beginning of the trial, followed by an interval of more stable durations, and a slight decrease towards the end of the trial. This pattern of change was found across both tasks and all path angle conditions. In parallel, the saccade amplitude gradually decreased in the LR-task and the PA-task when a contour was perceived, indicating that observers switched from a global and coarse-scale to a more local and fine-scale search strategy, presumably when observers encountered a likely contour candidate. This interpretation is supported by the observation that the mean distance of fixation to the target contour showed a similar decay over time when this contour was correctly detected. The time course of saccade amplitude is in line with a number of studies showing that information about the gist of the scene can be processed very fast after the onset of a stimulus and can guide initial saccades to regions where the target is likely to be found, so that these regions can be further explored using smaller saccades (Eckstein, Drescher, & Shimozaki, 2006; Neider & Zelinsky, 2006; Torralba, Oliva, Castelhano, & Henderson, 2006). It should be noted that in the LR-task, saccade amplitude first slightly increased during the first fixations. A possible reason is that observers initially make saccades to locations close to the fixation symbol present prior to stimulus presentation and only move towards more peripheral locations after a few fixations.

The decrease of saccade amplitude occurring together with an increase of fixation duration over time seems to reflect a general coarse-to-fine search strategy that has been frequently reported in earlier studies (Antes, 1974; Castelhano, Mack, & Henderson, 2009; Over et al., 2007; Pannasch et al., 2008; Scinto, Pillalamarri, & Karsh, 1986; Unema et al., 2005). The gradual decrease in saccade amplitudes, however, was not found when observers reported not perceiving a contour in the PA-task. In these trials, observers continued to make large saccades throughout the course of the trial. Thus, a coarse-to-fine search

process only seems to occur when a target candidate is found at the end of the trial. The coarse-to-fine processing strategy therefore does not seem to be a default mechanism of the visual system, as has been suggested by others (Over et al., 2007; Pannasch et al., 2008; Unema et al., 2005). A possible reason for the discrepancy between our findings and previous studies reporting a coarse-to-fine time course of saccade amplitude during visual search (Castelhano et al., 2009; Mills et al., 2011; Over et al., 2007; Scinto et al., 1986) could be that in these earlier studies the search target was eventually found in the majority of the trials. Presumably observers made smaller saccades near the end of a trial to process the image region in which the target was found in more detail. In contrast, in a visual search study in which the search target was rarely detected, observers made large saccades which remained relatively constant over the course of the trial (Mills et al., 2011). It was argued that this result was due to the fact that the images contained no contextual information that could guide participants' search behavior. Observers might have therefore adopted a strategy of making large saccades to visit as many locations as possible, since more in-depth processing of the scene did not help in finding the target location. In our study, it is likely that for trials in which observers did not perceive a contour there were no image regions that attracted their attention for further scrutiny, resulting in observers using a similar strategy as in the study of Mills et al. (2011). Consistent with this idea, the small increase in saccade amplitude for trials in which observers reported not perceiving a contour can be predicted by a random search model that visits stimulus locations which have not been investigated before until the whole visual field has been searched (simulations not shown). In consequence, the slight increase in saccade amplitude over time probably does not reflect an active adaptation of the search strategy, but is rather caused by geometrical constraints. The notion of a rather random search is also supported by the lower predictive power ($R = 0.583$) of the model for contour-absent stimuli than for stimuli in which a contour is present ($R = 0.64$).

It is important to emphasize that the dynamics of saccade amplitudes in the PA-task did not depend so much on the actual stimulus content (contour present or absent), but rather on the percept of observers (contour perceived or not perceived). A similar difference in the dynamics of saccade amplitudes was not found in the LR-task. This is presumably

due to the fact that there was always a contour present in the LR-task, causing observers to be “forced” to find a contour. As a consequence, they are more likely to “home in” on a potential target even if it is not the contour that was originally hidden in the stimulus. In the PA-task, on the other hand, observers are “allowed” to make the decision that no contour is present, resulting in trials in which no detailed processing of particular image regions takes place and long saccades prevail. Overall, saccade amplitudes were larger in the LR-task than in the PA-task. A possible reason is the existence of a division zone at the center of the display in the LR-task. Observers do not expect contours in this zone and therefore seldom make eye fixations towards it. This mental division of the display in two hemifields may have caused larger saccade amplitudes when observers jump between hemifields.

In sum, fixation durations appear to be determined by the saliency of a present contour from the beginning of the trial and systematically increase over the course of the trial, resulting in more detailed processing. In contrast, the time course of saccade amplitudes depends on the potential locations of the target contour and whether or not the contour can be perceived, irrespective of the saliency of a present contour. These results imply that fixation duration and saccade amplitude are not controlled by a common mechanism, as has been suggested by others (Pannasch et al., 2008; Unema et al., 2005), but instead that task and stimulus factors influence these parameters differently. A similar view has been put forward by Mills et al. (2011), arguing that the increase of fixation duration reflects a gradual built-up towards a more complex representation of the image, whereas the pattern of change in saccade amplitude depends on the utility of visual information available in the display.

2.4.3 Predictive power of contour integration model

The contour integration model was calibrated and tested on independent data. For determining model parameters, we used the decisions of observers in the LR-task (deciding whether the contour was in the left or right hemifield). Based on previous findings showing that the range of parameter values of the association field reaching or exceeding

mean human performance is very broad (Ernst et al., 2012; Watt et al., 2008), the model was required to also reproduce correlations in human contour detection behavior rather than only mean human performance, which resulted in a better parameter selection criterion. The predictive power of the model was then quantified on the eye tracking data, where we explicitly excluded eye movements in the vicinity of the target contour (which are indicative of all correct decisions made). The model was able to predict observers' fixation locations well above chance level in both the LR-task and the PA-task. As a consistency check, we compared model performances in reproducing behavioral decisions and in predicting eye movements. This analysis confirmed that the parameters found during model calibration are also optimal for predicting fixation locations. These results strengthen the validity of our model, as it not only reproduces human response behavior, but also predicts the salient image locations at which observers allocate their attention during contour integration. Hence, our modelling approach seems to be able to reconstruct the contour integration process which guides both perceptual decisions and eye movement behavior.

The ROC analysis and visual comparison of fixation locations with model predictions demonstrated that many, but not all fixations were predicted. The worst performance was actually found for the PA-task when contours were absent, which we attribute to a more random search strategy (see above). ROC integral values range between 0.58 and 0.64, which is in the same range as previously reported values from other models predicting fixations during image viewing (i.e., from 0.55 to .70; e.g., Betz, Kietzmann, Wilming, & König, 2010; Einhäuser, Spain, & Perona, 2008; Renninger, Verghese, & Coughlan, 2007). Importantly, such eye movement models include additional assumptions concerning the sequence of fixation locations (e.g., winner-take-all principle, inhibition of return to previously visited fixation locations). Given that our model is based on a simple association field computation producing one saliency map, performance is surprisingly good compared to that of more complex models.

Moreover, the model might appear worse than it actually is because the comparison to the real data involves several technical difficulties. First, fixations often do not go directly to a target, but instead first go into the vicinity of a (putative) contour (about

2-3 degrees of visual angle). Thus, the comparison radius of model “hotspots” to fixations must be large, causing also the number of false positives in the surrogate data to increase. Second, model activity evolves over time and might predict different fixation locations at different times in the simulation. The reason is the recurrent nature of the contour integration process: over time, longer and longer contour candidates are found, while shorter segments decrease in saliency. Nevertheless, even in the PA-task for stimuli without a contour, the correlation between model predictions and experimental data was much higher than chance level. This observation was confirmed by computing the distribution of correlations between saliency and fixation maps, which has a distance of $d' \approx 2$ to surrogate data. It is important to note that our goal was not to predict the temporal sequence of fixations during a trial, as this would require a more complex model that takes into account a number of assumptions concerning the planning and monitoring of subsequent eye movements (see Itti & Koch, 2001; Torralba et al., 2006; Zelinsky & Sheinberg, 1997 for examples of computational eye movement models). Rather, we aimed to test whether a model providing a saliency map based on a single association field is able to predict both contour detection performance as well as the locations likely to be fixated over the entire course of a trial.

2.4.4 Differences between observers

There are strong inter-individual differences between observers both in the experimental data (e.g., number of fixations made), and in the model’s predictive power. These differences are consistent across data sets (i.e., LR-task and PA-task). Another prominent finding is a negative correlation between the mean number of fixations an observer makes and the predictive power of the model. A possible explanation is that observers whose data is not well predicted by the model tend to make eye movements not towards putative contour candidates with high saliency, but to more “random” locations in the visual field. Interestingly, the mean number of fixations also correlated with the performance of observers (slopes 0.014 and 0.009 for a linear fit with $R^2 = 0.16$ and $R^2 = 0.07$ for the LR-task and PA-task, respectively, data not shown). These dependencies suggest that making more eye movements, including less salient locations in the visual field as well, is a

more successful strategy to find contour targets. A further difference between individual observers is the model’s integration time needed until the best match to their behavior is reached. This difference might indicate that observers differ in their criteria for contour candidates, i.e., they may require different lengths of aligned edge configurations to be salient and to be used as targets for a saccade.

2.4.5 Saccades compensate for reduced saliency in peripheral vision

In our procedure to establish the best parameters for the contour integration model, we also allowed for an optional scaling of edge saliency with eccentricity. The inclusion of this option was based on several visual search studies who have reported that detection performance of a target declines at higher eccentricities due to loss of spatial resolution (Carrasco, Evert, Chang, & Katz, 1995; Carrasco & Frieder, 1997; Carrasco & Yeshurun, 1998; Geisler & Chou, 1995; Geisler, Perry, & Najemnik, 2006; Scialfa & Joffe, 1998). In addition, the crowding literature shows that it is particularly difficult to identify a target in the periphery when it is surrounded by nearby distractor stimuli (Bouma, 1970; Levi, 2008). Eye movements help to explore these peripheral target locations and bring them to the foveal region, thereby reducing crowding and allowing for detailed spatial processing of these locations (Harrison, Mattingley, & Remington, 2013; Vlaskamp & Hooge, 2006; Wertheim, Hooge, Krikke, & Johnson, 2006). This latter observation is in agreement with our finding that contour detection behavior was best explained when scaling of saliency with eccentricity was absent, indicating that observers’ eye movements can fully compensate for the reduced contour saliency in peripheral vision.

2.4.6 Conclusions

A crucial task of everyday vision involves the grouping of spatially separate elements into more global structures such as contours. Eye movements play an important role in the detection of contours, allowing observers to fixate elements which belong to potential contours. The aim of our study was to examine how spatial and temporal aspects of sac-

saccadic eye movements interact with the process of contour integration. A relatively simple model of contour integration, based on an association field with parameters that best matched human contour integration performance, could successfully predict a considerable amount of fixation locations before the observers found the contour. This indicates that the observers' fixations are drawn to collinear structures, determined by the saliency map of the model. Model fitting further revealed that saccadic eye movements equilibrate the saliency of peripheral contour candidates across the visual field. As such, our study provides a more stringent test of the validity of the model by requiring it to predict the locations where observers fixate during a contour integration task, rather than requiring it to merely reproduce contour detection performance. Moreover, we find that observers' eye movements are adjusted to the difficulty of the contour integration task: temporal and spatial oculomotor parameters such as individual fixation durations and saccade amplitudes follow a specific time course that is strategically adjusted to the saliency and the percept of a contour in a given trial. The latter findings have important implications for current computational models of eye movement behavior during visual search. These models are typically used to predict the positions of fixations, but only incorporate dynamical aspects of eye movements to a limited extent. Our study shows that spatial and temporal aspects of eye movements can be used to advance our understanding of visual integration processes by providing rich datasets that allow for the refinement of computational models of these processes.

References

- Antes, J. R. (1974). The time course of picture viewing. *Journal of Experimental Psychology*, 103(1), 62–70.
- Betz, T., Kietzmann, T. C., Wilming, N., & König, P. (2010). Investigating task-dependent top-down effects on overt visual attention. *Journal of Vision*, 10(3), 1–14.
- Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature*, 226(5241), 177–178.
- Carrasco, M., Evert, D. L., Chang, I., & Katz, S. M. (1995). The eccentricity effect: Target eccentricity affects performance on conjunction searches. *Attention, Perception, & Psychophysics*, 57(8), 1241–1261.
- Carrasco, M., & Frieder, K. S. (1997). Cortical magnification neutralizes the eccentricity effect in visual search. *Vision Research*, 37(1), 63–82.
- Carrasco, M., & Yeshurun, Y. (1998). The contribution of covert attention to the set-size and eccentricity effects in visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 24(2), 673–692.
- Castelhano, M. S., Mack, M. L., & Henderson, J. M. (2009). Viewing task influences eye movement control during active scene perception. *Journal of Vision*, 9(3), 1–15.
- Eckstein, M. P., Drescher, B. A., & Shimozaki, S. S. (2006). Attentional cues in real scenes, saccadic targeting, and Bayesian priors. *Psychological Science*, 17(11), 973–980.
- Einhäuser, W., Spain, M., & Perona, P. (2008). Objects predict fixations better than early saliency. *Journal of Vision*, 8(14), 1–26.
- Ernst, U. A., Mandon, S., Schinkel-Bielefeld, N., Neitzel, S. D., Kreiter, A. K., & Pawelzik, K. R. (2012). Optimality of human contour integration. *PLoS Computational Biology*, 8(5), e1002520.
- Field, D. J., Hayes, A., & Hess, R. F. (1993). Contour integration by the human visual system: evidence for a local “association field”. *Vision Research*, 33(2), 173–193.
- Foley, J. M., Varadharajan, S., Koh, C. C., & Farias, M. C. Q. (2007). Detection of Gabor patterns of different sizes, shapes, phases and eccentricities. *Vision Research*, 47(1), 85–107.
- Geisler, W. S., & Chou, K.-L. (1995). Separation of low-level and high-level factors in complex tasks: visual search. *Psychological Review*, 102(2), 356–378.
- Geisler, W. S., Perry, J. S., & Najemnik, J. (2006). Visual search: The role of peripheral information measured using gaze-contingent displays. *Journal of Vision*, 6(9), 858–873.
- Harrison, W. J., Mattingley, J. B., & Remington, R. W. (2013). Eye movement targets are released from visual crowding. *Journal of Neuroscience*, 33(7), 2927–2933.
- Henderson, J. M., & Pierce, G. L. (2008). Eye movements during scene viewing: Evidence for mixed control of fixation durations. *Psychonomic Bulletin & Review*, 15(3), 566–573.
- Henderson, J. M., & Smith, T. J. (2009). How are eye fixation durations controlled during scene viewing? Further evidence from a scene onset delay paradigm. *Visual Cognition*, 17(6–7), 1055–1082.

- Hess, R. F., & Dakin, S. C. (1997). Absence of contour linking in peripheral vision. *Nature*, 390(6660), 602–604.
- Hess, R. F., May, K. A., & Dumoulin, S. O. (2013). Contour integration: psychophysical, neurophysiological and computational perspectives. In J. Wagemans (Ed.), *Oxford handbook of perceptual organization*. Oxford, UK: Oxford University Press.
- Hoffman, L., & Rovine, M. J. (2007). Multilevel models for the experimental psychologist: Foundations and illustrative examples. *Behavior Research Methods*, 39(1), 101–117.
- Hooge, I. T. C., & Erkelens, C. J. (1996). Control of fixation duration in a simple search task. *Attention, Perception, & Psychophysics*, 58(7), 969–976.
- Itti, L., & Koch, C. (2001). Computational modeling of visual attention. *Nature Reviews Neuroscience*, 2(3), 194–203.
- Jacobs, A. M. (1986). Eye-movement control in visual search: how direct is visual span control? *Attention, Perception, & Psychophysics*, 39(1), 47–58.
- Levi, D. M. (2008). Crowding—An essential bottleneck for object recognition: A mini-review. *Vision Research*, 48(5), 635–654.
- Li, Z. (1998). A neural model of contour integration in the primary visual cortex. *Neural Computation*, 10(4), 903–940.
- Locke, L., Hoffman, L., & Bovaird, J. A. (2007). On the use of multilevel modeling as an alternative to items analysis in psycholinguistic research. *Behavior Research Methods*, 39(4), 723–730.
- May, K. A., & Hess, R. F. (2007). Ladder contours are undetectable in the periphery: A crowding effect? *Journal of Vision*, 7(13), 1–15.
- May, K. A., & Hess, R. F. (2008). Effects of element separation and carrier wavelength on detection of snakes and ladders: Implications for models of contour integration. *Journal of Vision*, 8(13), 1–23.
- McCarley, J. S., Kramer, A. F., Boot, W. R., Peterson, M. S., Wang, R. F., & Irwin, D. E. (2006). Oculomotor behaviour in visual search for multiple targets. *Visual Cognition*, 14(4–8), 685–703.
- Mills, M., Hollingworth, A., Van der Stigchel, S., Hoffman, L., & Dodd, M. D. (2011). Examining the influence of task set on eye movements and fixations. *Journal of Vision*, 11(8), 1–15.
- Mundhenk, T. N., & Itti, L. (2002). A model of contour integration in early visual cortex. In *Second international workshop on biologically motivated computer vision* (pp. 80–89). Heidelberg: Springer.
- Neider, M. B., & Zelinsky, G. J. (2006). Scene context guides eye movements during visual search. *Vision Research*, 46(5), 614–621.
- Nugent, A. K., Keswani, R. N., Woods, R. L., & Peli, E. (2003). Contour integration in peripheral vision reduces gradually with eccentricity. *Vision Research*, 43(23), 2427–2437.
- Over, E. A. B., Hooge, I. T. C., Vlaskamp, B. N. S., & Erkelens, C. J. (2007). Coarse-to-fine eye movement strategy in visual search. *Vision Research*, 47(17), 2272–2280.
- Pannasch, S., Helmert, J. R., Roth, K., Herbold, A. K., & Walter, H. (2008). Visual fixation durations and saccade amplitudes: Shifting relationship in a variety of conditions. *Journal of Eye Movement Research*, 2(2), 1–19.

- Renninger, L. W., Verghese, P., & Coughlan, J. (2007). Where to look next? Eye movements reduce local uncertainty. *Journal of Vision*, 7(3), 1–17.
- Scialfa, C. T., & Joffe, K. M. (1998). Response times and eye movements in feature and conjunction search as a function of target eccentricity. *Attention, Perception, & Psychophysics*, 60(6), 1067–1082.
- Scinto, L. F. M., Pillalamarri, R., & Karsh, R. (1986). Cognitive strategies for visual search. *Acta Psychologica*, 62(3), 263–292.
- Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford: Oxford University Press.
- Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. *Psychological Review*, 113(4), 766–786.
- Unema, P. J. A., Pannasch, S., Joos, M., & Velichkovsky, B. M. (2005). Time course of information processing during scene perception: The relationship between saccade amplitude and fixation duration. *Visual Cognition*, 12(3), 473–494.
- Velichkovsky, B. M., Joos, M., Helmert, J. R., & Pannasch, S. (2005). Two visual systems and their eye movements: Evidence from static and dynamic scene perception. In B. G. Bara, L. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the xxvii conference of the cognitive science society* (pp. 2283–2288). Mahwah, NJ.
- Vlaskamp, B. N. S., & Hooge, I. T. C. (2006). Crowding degrades saccadic search performance. *Vision Research*, 46(3), 417–425.
- Vlaskamp, B. N. S., Over, E. A. B., & Hooge, I. T. C. (2005). Saccadic search performance: the effect of element spacing. *Experimental Brain Research*, 167(2), 246–259.
- Wagemans, J., Elder, J., Kubovy, M., Palmer, S., Peterson, M., Singh, M., & von der Heydt, R. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure-ground organization. *Psychological Bulletin*, 138(6), 1172–1217.
- Watt, R., Ledgeway, T., & Dakin, S. C. (2008). Families of models for Gabor paths demonstrate the importance of spatial adjacency. *Journal of Vision*, 8(7), 1–19.
- Wertheim, A. H., Hooge, I. T. C., Krikke, K., & Johnson, A. (2006). How important is lateral masking in visual search? *Experimental Brain Research*, 170(3), 387–402.
- Williams, L. R., & Thornber, K. K. (2001). Orientation, scale, and discontinuity as emergent properties of illusory contour shape. *Neural Computation*, 13(8), 1683–1711.
- Yen, S.-C., & Finkel, L. H. (1998). Extraction of perceptually salient contours by striate cortical networks. *Vision Research*, 38(5), 719–741.
- Zelinsky, G. J., & Sheinberg, D. L. (1995). Why some search tasks take longer than others: Using eye movements to redefine reaction times. In R. K. J. Findlay & R. Walker (Eds.), *Eye movement research: Mechanisms, processes and applications* (Vol. 6, pp. 325–336). Amsterdam: Elsevier Science.
- Zelinsky, G. J., & Sheinberg, D. L. (1997). Eye movements during parallel–serial visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 23(1), 244–262.

Chapter 3

Presaccadic EEG activity reflects visual saliency during contour integration

This chapter is a manuscript in preparation:

Van Humbeeck, N., Nikolaev, A. R., van Leeuwen, C., & Wageman, J. (2016). *Presaccadic EEG activity reflects visual saliency during contour integration*. Manuscript in preparation.

3.1 Introduction

When searching for an object in a cluttered scene, we perform saccadic eye movements in order to bring relevant locations into the foveal region. Saccades are controlled by competing top-down and bottom-up processes (Kristjánsson, 2007). Top-down factors that play an key role in guiding eye movements are based on prior knowledge, intentions and task goals (Henderson, 2003; Rothkopf, Ballard, & Hayhoe, 2007; Tatler, Hayhoe, Land, & Ballard, 2011; Torralba, Oliva, Castelhano, & Henderson, 2006; Wolfe, Butcher, Lee, & Hyle, 2003); bottom-up factors involve visual saliency of stimulus features which may automatically “pop-out” from the background. In the extreme case, identification of a peripheral target occurs prior to a saccade; the following fixation merely confirms target detection (Kotowicz, Rutishauser, & Koch, 2010). Based on bottom-up salient information, computational models perform considerably better than chance in predicting fixation locations in a scene (Foulsham & Underwood, 2008; Itti & Koch, 2001; Koch & Ullman, 1985; Parkhurst, Law, & Niebur, 2002), suggesting that visual saliency contributes to saccade guidance.

In Chapter 2, we examined the role of visual saliency in saccade guidance during a free-viewing contour integration task. Participants identified a global contour hidden in a background of randomly oriented Gabor patches. We proposed a contour integration model based on an association field, in which regions containing high collinear structure, as manifested in high “association strength”, received high scores on a saliency map. The map predicted fixation locations significantly above chance level. We additionally found that even at the beginning of a free-viewing trial, an initial but still incomplete saliency map of the search display seems to exist and effectively guides saccade planning.

Here we aim to study the brain processes involved in fixation location selection in a contour integration task. For this, we will use co-registration of electroencephalography (EEG) and eye movements. Saccade-related and fixation-related EEG activity has been measured during search tasks in which certain target stimuli are presented among distractors (Kelly, Foxe, Newman, & Edelman, 2010; Mazaheri, DiQuattro, Bengson, & Geng, 2011; Ptak, Camen, Morand, & Schnider, 2011). It has been found that, when observers

manage to initiate saccades to the targets rather than to the distractors, a specific pattern of EEG amplitude can be measured during the interval prior to the onset of the saccade. The presaccadic EEG amplitude pattern is different when saccades are directed to the distractors. A second observation is that EEG activity in the presaccadic interval indicates the amount of covert attention that is deployed to the next fixation location (Gutteling, van Ettinger-Veenstra, Kenemans, & Neggers, 2010; Krebs, Boehler, Zhang, Schoenfeld, & Woldorff, 2012; Wauschkuhn et al., 1998). In summary, fixating on a relevant location (e.g., a target in a search task) and ignoring irrelevant locations seems to require specific brain processes operating during the presaccadic interval. These processes lead to specific patterns of EEG activity. Presumably, this presaccadic EEG activity reflects the level of attention required to ignore irrelevant locations, e.g., to suppress fixations on distractors in a search task.

In the context of a contour detection task, relevant locations are those associated with high association strength. The results of Chapter 2 suggest that high association strength is associated with high saliency and that saliency guides saccade planning. If this is indeed the case, we expect saliency to evoke the aforementioned presaccadic attentional brain processes and the corresponding pattern of EEG activity. More specifically, fixations on locations of high association strength (i.e., relevant, salient locations) are expected to evoke a different pattern of presaccadic EEG activity compared to fixations on locations of low salience. In other words, presaccadic EEG activity should be predictive of visual saliency at the fixation location.

In EEG-eye movement co-registration research, saccades are natural markers for segmenting behavior and related brain activity. Saccade-based segmentation is most advantageous in free-viewing behavior, which lacks external segmentation markers such as stimulus onset or response initiation. An increasing number of studies of visual search have used EEG-eye movement co-registration in natural viewing conditions (e.g., Devillez, Guyader, & Guérin-Dugué, 2015; Kamienkowski, Ison, Quiroga, & Sigman, 2012; Kau-nitz et al., 2014; Wenzel, Golenia, & Blankertz, 2016). However, only a handful of them investigated EEG activity prior to saccade onset (Dandekar, Privitera, Carney, & Klein, 2012; Dias, Sajda, Dmochowski, & Parra, 2013; Körner et al., 2014). In the study of Dias

et al. (2013), a known target had to be detected among distractors that shared features, such as color or orientation, with the target. As long as these distractors did not make the task too difficult, presaccadic activity as early as 150 ms before saccade onset could predict whether a saccade was landing on a distractor or a target. This effect was found only in conditions where manual choice responses were required, suggesting that the presaccadic activity reflected preparation of the manual response. However, presaccadic activity could also reflect selection of the next fixation target in the absence of any motor response. When memorizing a natural scene in preparation for a subsequent change detection task, presaccadic potentials at parietal and occipital sites predicted whether a saccade would land on a prominent location (Nikolaev, Jurica, Nakatani, Plomp, & van Leeuwen, 2013). The prominence was determined according to the average duration and density of observers' fixations. In other studies, EEG potentials time-locked to saccade and fixation onsets indicated that prior detection of a target modulates processing of distractors on subsequent fixations (Körner et al., 2014). These findings may illustrate that co-registration can reveal the effects of saliency or task relevance on presaccadic EEG activity in free-viewing conditions.

Caution is needed in interpreting brain signals in free-viewing behavior, since eye movements induce large artifacts on the EEG signal. These artifacts are not limited to the rotation of the corneo-retinal dipole and constriction of ocular muscles, of which the effects on EEG amplitude are known from classical studies using stimulus-evoked potentials (ERP) (Gratton, Coles, & Donchin, 1983; Jung, Makeig, Humphries, et al., 2000; Lins, Picton, Berg, & Scherg, 1993). As opposed to ERP research, the main problem in analyzing the EEG in free-viewing is that neural responses to sequential eye movements may overlap. This makes it difficult to isolate neural activity associated with a single saccade, since EEG responses to temporally adjacent saccades contaminates the EEG response to the current saccade (Dandekar et al., 2012; Dimigen, Sommer, Hohlfeld, Jacobs, & Kliegl, 2011; Nikolaev, Meghanathan, & van Leeuwen, 2016). A common approach to address this issue has been to use trial selection procedures for matching eye movements on a number of criteria, such as fixation duration and saccade size, between experimental conditions (e.g., Brouwer, Reuderink, Vincent, van Gerven, & van Erp, 2013; Dias et al.,

2013; Kamienkowski et al., 2012). However, even with a limited number of matching criteria, matching is likely to result in a significant loss of data. This not only reduces the statistical power of the analysis but may also eliminate processes of interest that are associated with the removed (i.e., unmatched) data.

Another strategy that has been used for correcting the effects of nearby saccades involves linear regression techniques. These are applied to estimate the neural response to a saccade while partialling out the influence of temporally close eye movements (Dandekar et al., 2012; Dimigen et al., 2011). However, these techniques are insensitive to nonlinear influences of overlapping EEG responses (Dimigen, 2014). Given the nonlinearity of the brain signal (Cohen, 2014), these regression methods are likely to miss a large proportion of the contamination. In particular, the duration of a fixation modulates the temporal onset of the lambda response, a positive wave peaking around 100 ms after the fixation onset. Consequently, any variation in fixation duration will have a nonlinear influence on the brain activity during the presaccadic interval. This effect will interact with the size of the saccade occurring before this fixation, as it is known that lambda responses increase with saccade size (Thickbroom, Knezevic, Carroll, & Mastaglia, 1991; Yagi, 1979).

In the current study, we will implement a method to deal with the nonlinear effect of overlapping EEG responses. We will apply a statistical model that partials out the effects of confounding variables on the EEG response by including them as covariates. Specifically, we will apply Generalized Additive Mixed Modelling (GAMM, Hastie & Tibshirani, 1990; Tremblay & Newman, 2015; Wood, 2006), developed to model nonlinear relationships between predictors and the response variable. This method models the EEG signal as an additive structure of nonlinear smoothing functions. The method enables us to disentangle the EEG activity associated with eye movement characteristics from those related to the experimental event of interest. Moreover, the GAMM approach for EEG analysis has several additional advantages (Meulman, Wieling, Sprenger, Stowe, & Schmid, 2015; see also Vossen, Van Breukelen, Hermens, Van Os, & Lousberg, 2011 showing the suitability of mixed-effects models in general for EEG analysis). First, EEG signals are analysed on a single-trial basis and subsequently trial-to-trial variance within participants can be modelled. The single-trial analysis is advantageous in comparison to

the traditional averaging methods commonly used in EEG research, because stimulus- and task-relevant information may be lost during trial averaging (Alexander et al., 2013). Second, GAMM allows for the analysis of the (non-linear) time course of the EEG signal as well as latency differences in the effects of time-varying predictors. Third, GAMM can incorporate random effects, and hence is ideally suited to deal with item and participant variation. Finally, GAMM can deal with the analysis of unbalanced data, a situation which commonly occurs in EEG research.

The goal of this study is to explore the attentional brain processes involved in the presaccadic selection of a target in a free-viewing contour detection task. More specifically, we test the hypothesis that presaccadic EEG activity is predictive of the visual saliency of the fixation location. To this end, we will present participants with dense search displays containing randomly oriented Gabor elements in which a smooth target contour was either present or not (similar to the present-absent task in Chapter 2). As mentioned earlier, we predict that fixations on locations of high association strength will evoke a different pattern of presaccadic EEG activity compared to fixations on other locations. In that case, presaccadic EEG activity can indeed be considered predictive of visual saliency at the fixation location and may reflect saliency-induced attentional processes. To test this prediction for contour-present trials, we measured the distance between the landing position of the upcoming saccade and the target contour. Indeed, the location of the contour is a salient location with high association strength. For contour-absent trials, we determined the association strength value at the fixation location (i.e., the saccadic end point) using an association field model. We measured EEG activity during each presaccadic interval and attempted to find a link between this activity and either the distance measure (for contour-present trials) or the association strength value (for contour-absent trials). We found evidence for such a link. Presaccadic EEG activity was found to be larger, especially over parietal and occipital brain areas, both when the distance between the fixation location and the contour was larger (in contour-present trials) as well as when the association strength value was lower (in contour-absent trials). In other words, large presaccadic EEG activity predicts a saccade towards a low salient location.

3.2 Methods

3.2.1 Participants

Twenty-three healthy adults (two male, mean age = 21.8, range = 18-33 years) with normal or correct-to-normal vision participated in the experiment in exchange for course credits or payment. All participants gave written consent before the start of the experiment. The experiment was approved by the Social and Societal Ethics Committee of the University of Leuven. Data from one participant was removed from further analysis due to accuracy errors during eye movement recording. Another participant was excluded because of EEG artifacts, as we will explain below.

3.2.2 Stimuli

The stimulus display consisted of an array ($30^\circ \times 30^\circ$) of Gabor patches. These were defined as the product of a cosine grating (spatial frequency of 3.7 cycles/ $^\circ$) and a Gaussian contrast envelope ($SD = 0.14^\circ$). Depending on presence of a curved contour rendered by aligning Gabor patches the stimuli were divided in two conditions.

In the contour-present condition, the stimulus display included a contour embedded in a background of randomly oriented Gabor elements. This contour was a sequence of seven Gabor patches with an angle of 25° between two successive elements. The sign of the angle formed between successive elements could be positive or negative with equal probability. The distance between the centers of neighboring elements in the contour was 0.7° . The rest of the stimulus display was filled with randomly oriented Gabor patches, positioned in such a way that the mean distance between neighboring elements in the background was also 0.7° .

In the contour-absent condition, the stimulus display consisted only of Gabor elements with random orientations, placed with the same mean distance of 0.7° between the elements. To find locations having collinear elements in this condition, we applied an association field model which calculated an association strength for each oriented ele-

ment, determining whether a particular element is likely to belong to a contour (Ernst et al., 2012; Field, Hayes, & Hess, 1993; Watt, Ledgeway, & Dakin, 2008). The association strength value is high when adjacent elements are co-circular and co-aligned along a smooth contour. Here, we used a simplified version of the association field model used in our previous study (see Chapter 2; Van Humbeeck, Schmitt, Hermens, Wagemans, & Ernst, 2013), which only depends on three parameters, namely the sensitivity to curvature σ_α , the sensitivity to co-circularity σ_β and the scaling constant λ . We used the same parameter values which were found to be optimal to match or to exceed human performance, i.e., $\sigma_\alpha = 0.15$, $\sigma_\beta = 0.24$, and $\lambda = 0.3$ (Ernst et al., 2012; see also Chapter 2).

In Chapter 2, we proved the validity of the association field model by assessing whether it could predict participants' fixation locations, even when fixations that were located close to the actual contour were excluded. More specifically, we evaluated whether fixations were located on "hotspots", i.e., regions with high associations strength, considerably more than chance level by performing a receiver-operator characteristics (ROC) analysis. The area under the ROC function, the ROC integral, served as a quantitative measure of the model's predictive power (for more details about this procedure see Methods section of Chapter 2; Van Humbeeck et al., 2013). We applied the same ROC analysis to examine the predictive power of the association field model used in the current study. If eye movements were guided by bottom-up salient locations (i.e., collinear structures) to some extent, our model should be able to predict fixation locations at a better than chance level.

To obtain a measure of association strength at each fixation positions, for each image we constructed heat maps which indicated the normalized association strength value at each pixel of the stimulus display. This was done by multiplying the original association strength values of the Gabor elements with a 2-dimensional Gaussian function ($SD = 1.2^\circ$) positioned at the center of the Gabor elements. The heat maps were normalized by dividing the association strength values obtained at each pixel by the sum of all pixel values in the map.

We constructed the stimuli using the GERT toolbox for MATLAB (Demeyer & Machilsen,

2012). This toolbox includes a procedure which eliminates local density cues of contour integration by controlling the spacing between elements. In such a way we ensured that only relative orientation cues were used for contour integration.

3.2.3 Procedure

Participants were seated in a dimly lit room at a distance of 55 cm from the screen. A chin rest reduced head movement. At the beginning of each trial, participants were asked to fixate a centrally presented fixation cross for a duration randomly selected between 500 and 1000 ms. The Gabor field stimulus was then presented for 8 s. During this period, participants were allowed to freely move their eyes across the search display. In the following response interval participants indicated whether a contour was present or not by pressing “p” or “q” keys (counterbalanced across participants) on a standard QWERTY keyboard. During this response interval, a question mark was shown in the center of the screen. If no response was made within a 5-s interval, no response was registered and a warning message prompted participants to respond faster. Feedback at the end of each trial indicated whether the response was correct or incorrect (see Figure 3.1 for an example of a trial sequence). Each participant was presented with 120 contour-present and 120 contour-absent Gabor fields, intermixed in random order, with 2 min breaks after each block of 40 trials. To familiarize participants with stimuli and task, a practice block containing five contour-present and five contour-absent trials was given before the main experiment.

3.2.4 Eye movement recording

Eye movements were recorded using a video-based infrared eye tracker (EyeLink 1000, desktop version, SR Research Ltd., Ontario, Canada). The movements of the right eye were tracked with a sampling frequency of 250 Hz. Saccades were detected using the default settings of the Eyelink software, using a velocity threshold of $30^{\circ}/s$ and an acceleration threshold of $8000^{\circ}/s^2$. Calibration and validation procedures were performed before the experiment and before each block. During the calibration procedure, partic-

ipants were instructed to fixate a dot presented at each of nine locations of a standard calibration screen. This procedure was repeated until their recorded fixations were aligned on a three by three rectangular grid. Before each trial, the participants were asked to fixate a dot in the center of the screen and to press the space bar. The keypress triggered correction for drifts in eye movement due to small head movements. If deviation between eye position and the central dot was larger than 2° , the eye tracker was recalibrated.

3.2.5 EEG recording

EEG was recorded using a 256-channel Electrical Geodesics System (Electrical Geodesics Inc., Eugene, OR). The electrode montage included sensors for recording vertical and horizontal electrooculogram (VEOG and HEOG). The recording was done at a sampling frequency of 250 Hz with an analog high-pass filter of 0.1 Hz and a 100-Hz low-pass filter. The recording reference was Cz. To synchronize stimulus presentation, EEG, and eye tracking signals, at the beginning of each trial a TTL pulse was sent via a Y-shaped cable

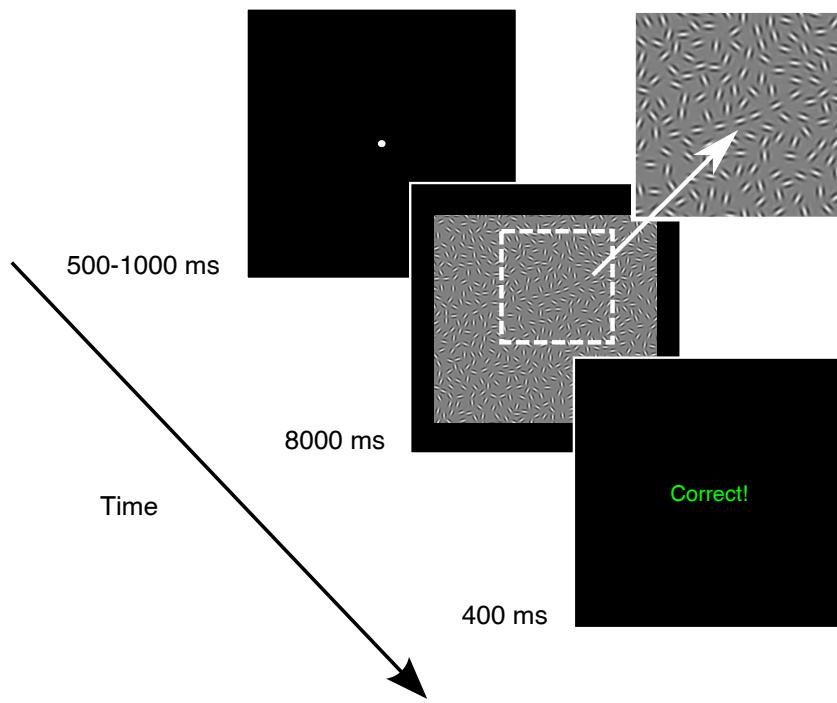


Figure 3.1. Illustration of the stimulus sequence. A fixation dot was presented for a random interval between 500 and 1000 ms. The stimulus display was presented until response but no longer than 8000 ms. Next, feedback was provided for 400 ms. The inset shows a close-up of the contour consisting of seven collinear elements.

attached to the parallel port of the stimulus presentation computer to both eye tracking and EEG recording systems.

3.2.6 Eye movement analysis

Both the eye movement and EEG analyses were conducted only for trials with correct responses, ensuring that a target contour was eventually detected in the contour-present condition and that no “false” contour was detected in the contour-absent condition. All analyses were carried out separately for the contour-present and contour-absent trials. For the contour-present trials, we considered only the saccades before the first visit to the contour. This event was registered when a fixation was located within a radius of 2 degrees of visual angle around a contour. Saccades performed after finding the contour were not counted; they do not form part of the visual search and are hence irrelevant to the purpose of the study. Saccades either landing or starting at a position outside the search image were also removed. This excluded 0.9% of the saccades.

Fixations shorter than 200 ms and larger than 1000 ms were excluded from further analysis, resulting in 28 % excluded fixations. In addition, we combined fixation intervals if an intervening saccade was smaller than 0.1° (which occurred in 0.03% of cases). Fixations before and after such saccades were concatenated to form a new fixation with a duration which is the sum of the two old fixation durations and a position which is the average of the two old fixation positions, weighted by their fixation durations. When fixations preceded saccades larger than 0.1° but smaller than 0.5°, both fixations and saccades were removed, which involved 0.65% of cases.

3.2.7 EEG preprocessing

The EEG data were preprocessed offline using the BrainVision Analyzer 2.0 software (Brain Products, GmbH, Germany). EEG was filtered with a Butterworth zero-phase filter with a low cut-off frequency of 0.5 Hz and a high cut-off frequency of 30 Hz, 48 dB/oct. We excluded 99 electrodes close to the cheeks and neck, which were contaminated with

strong muscle and other artifacts. Then we visually inspected the remaining electrodes and removed ones that appeared to be noisy. Data from one participant who had more than 15 channels distorted by large artifacts was excluded from further analysis.

To remove ocular artifacts, we used Independent Component Analysis (ICA; Jung, Makeig, Westerfield, et al., 2000; Makeig, Bell, Jung, & Sejnowski, 1996) for each participant separately. First, a 400-s interval of the continuous EEG recording which contained blinks and eye movements was fed into ICA. Based on a correlative score between the activity of the ICA components and the VEOG/HEOG channels, ICA components were identified in which the sum of squared correlations with the HEOG or the VEOG channel exceeded 0.3. These “ocular” components were subsequently removed and the entire EEG signal was recreated without these components. Next, we segmented EEG signal into epochs from -200 to 100 ms relative to the onset of a saccade. To remove EEG epochs containing other artifacts, such as large body movements, face/neck muscle activity, poor electrode contact, etc., we used the following criteria: an epoch was rejected if the absolute voltage difference between two adjacent sampling points exceeded 50 μ V and if the amplitude difference within a time interval of 100 ms exceeded +100 or -100 μ V. On average, 5% of epochs per participant were rejected, providing 677 ($SD = 147$) kept epochs per participant for the contour-present trials and 1709 ($SD = 211$) epochs for the contour-absent trials.

Baseline correction was performed by subtracting from each EEG data point the mean amplitude in the interval from 200 to 180 ms before saccade onset. We selected eight regions of interest (ROIs) over the left and right frontal, central, parietal and occipital brain regions. The ROIs were defined around the landmark electrodes of the International 10-20 System: F3, F4, C3, C4, P3, P4, O1, O2. For each ROI the amplitudes over one central and six surrounding electrodes were averaged.

3.2.8 Generalized additive mixed modelling

In contrast to traditional ERP analysis we did not average across trials, but performed our further analyses on single EEG epochs in the interval from -180 to -20ms before the

onset of the saccade. A dataset was created with the EEG amplitude for each ROI (the left and right frontal, central, parietal and occipital regions) as dependent variable and the following predictors: Participant, Image, Trial Number, Time, Fixation Duration, Preceding Saccade Size, and Current Saccade Size (see Figure 3.2). The predictor Image represented the specific search display, while Trial Number denoted the index of a trial during the experiment. The variable Time represented the time within the presaccadic interval (in ms). The predictor Fixation Duration was introduced because in free-viewing it may affect the amplitude of the presaccadic EEG activity due to overlap with the preceding lambda activity (Dimigen et al., 2011; Nikolaev et al., 2016). Saccade Size was introduced because it is known to affect the lambda activity occurring in the presaccadic interval offset(Keren, Yuval-Greenberg, & Deouell, 2010; Yagi, 1979). The overlapping effects may propagate, in principle, not only from the preceding eye movement (n-1), but also from the earlier eye movements (e.g., n-2, n-3). We did not include the effects of earlier eye movements into the model. In view of their purported weakness, adding these effects would have rendered model selection disproportionately more complex .

Our main variables of interest were as follows. For the contour-present trials, it was Distance, referring to the distance between the landing position of the upcoming saccade and the target contour. For the contour-absent trials, it was Association Strength, referring to the association strength at the landing position of the upcoming saccade, as predicted by an association field model (see Figure 3.2 for a schematic illustration of these two predictors).

Data analysis was carried out using the statistical software R. We applied a GAMM analysis using the *mgcv* package 1.8-6 in R (Wood, 2015). GAMM is a relatively recent statistical tool to model nonlinear relationships between predictors and the dependent variable. The possibility to examine nonlinear relationships is critical for analyzing EEG amplitude because EEG exhibits nonlinear dynamical properties (Cohen, 2014). GAMM provides an extension to the Generalized Linear Mixed Model (GLMM) by combining a possible parametric part with a non-parametric part. The latter part provides nonlinear smoothing functions for modelling wiggly surfaces of two or more numerical predictors. Over- and undersmoothing is avoided by applying penalized regression to the smoothing

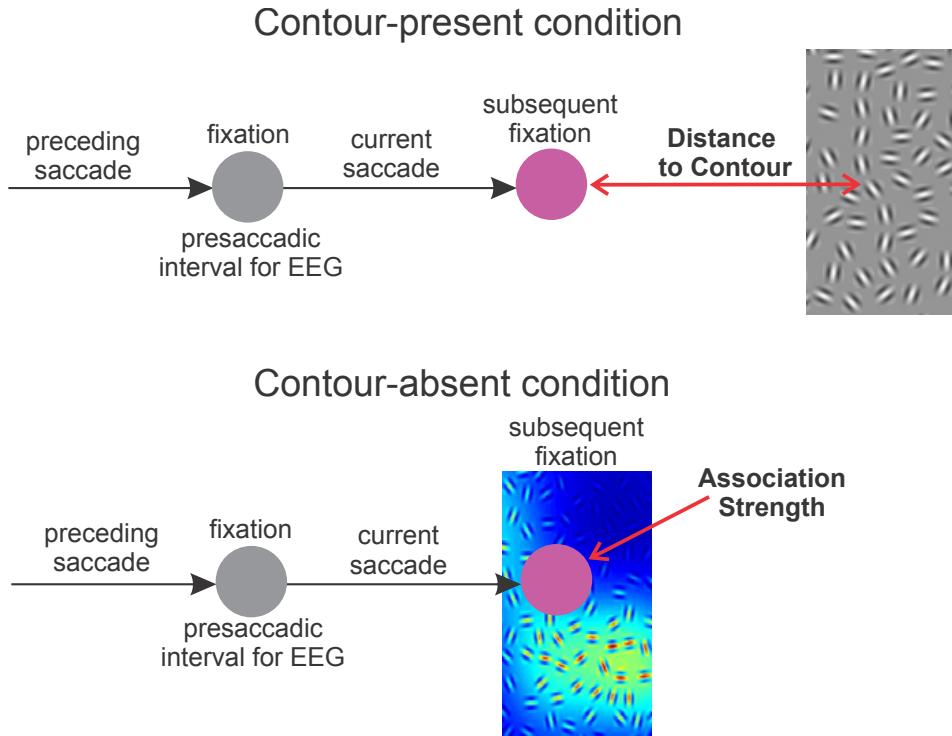


Figure 3.2. Definition of the two main variables of interest relative to the eye movement events. For the contour-present trials, the main variable is the distance between the landing position of the saccade and the target contour. For the contour-absent trials, the main variable is the association strength value at the saccade landing position.

functions (Hastie & Tibshirani, 1990; Tremblay & Newman, 2015). More specifically, a “wiggliness penalty” is added to the least squares fitting method. This penalty value λ is chosen by generalized cross-validation. When the penalty is high, the smoothing function will almost be a straight line. When it is close to zero, the smoothing function fit will almost entirely rely on the least-squares fitting method. In addition, since oversmoothing can still occur when the basis dimensions are chosen too small, the choice of basis dimensions used for the smooth terms are checked by a procedure provided by the *mgcv* package. Like GLMM, GAMM makes the distinction between fixed effects and random effects. In our data, the factors Participant and Image were treated as random-effect variables, as they could be considered a sample of a larger population of possible participants or images. Including these random-effects variables makes it possible to account for structural variation across participants and images. In addition, GAMM allows for the correction of autocorrelation present in the residual error. It is crucial for the analysis of EEG time series because its neighboring time points are highly correlated. Following

close observation of our error residuals, we included in our final model an autocorrelation parameter $\rho = 0.96$. As a result, almost no autocorrelation remained in the residual error of the model. These corrections were performed with the *itsadug* package (version 1.0.1; van Rij, Wieling, Baayen, & van Rijn, 2015) in R.

Using GAMM, we modelled an EEG signal as an additive structure of waveforms. We started from a simple model describing the general time pattern of EEG over participants and conditions and systematically added new predictors. Nonlinear effects of a single predictor are described by smooth terms, while nonlinear interactions between two predictors are described by tensor products (see Wood, 2006 for more details). For each added smooth or parametric term, we tested the significance of its contribution by comparing the fit of the extended model with the one not including the term. Model improvement was then examined by a Chi-Square test on two times the difference in maximum-likelihood scores and the difference in degrees of freedom. In addition, the p -values associated with the estimates of the parametric and smooth terms in the model were used to evaluate whether a predictor significantly contributed to the variance in EEG amplitude (for more information about model evaluation, see the vignettes in the *itsadug* package). Following this procedure, a final GAMM model was obtained in both the contour-present and contour-absent analysis.

To model the non-linear relationships between predictors and the EEG amplitude, we use thin plate regression splines for the smoothing functions as these are considered the best choice to estimate non-linearities and are computationally efficient (Meulman et al., 2015; Wood, 2003). The model included a smooth term for Time, representing the partial effect of Time on EEG amplitude. For our random-effect variable Participant, we included a nonlinear factor smooth for Time to model the presaccadic time course of EEG amplitude. In addition, we included a nonlinear factor smooth for Trial Number to model the change in EEG amplitude over the course of the experiment. For the random-effect variable Image, we examined whether there were systematic differences in average EEG amplitude across images by adding a random intercept for Image to the model. In addition, we tested the significance of smooth terms for the covariates Fixation Duration, and Preceding Saccade Size, and tensor products for the interaction of Time and

Fixation Duration, and Time and Preceding Saccade Size. For the contour-present trials, we examined the contribution of our main variable of interest, Distance to the contour, and its interaction with Time. For the contour-absent trials, Association Strength and its interaction with Time was considered. For all smooth and interaction terms (i.e., tensor products) except the random-effect variables, we assessed whether these effects differed for different brain regions of interest (represented by the parametric term ROI). Figure 3.3 illustrates the components of the GAMM model for the contour-present trials for a representative participant.

We also tested whether the saccade size could be predicted by the distance between the saccade landing position and the contour (in the contour-present trials) or the association strength at the saccade landing position (in the contour-absent trials). To this end, linear mixed modelling was applied with Participant and Image as random-effect variables and Distance to the contour and Association Strength as predictors of Saccade Size. The analyses were conducted using the *lme4* package in R (Bates, Mächler, Bolker, & Walker, 2015). Since histograms of Saccade Size and Association Strength showed skewed distributions, these variables were logarithmically transformed to approximate normal distributions (the same transformation was carried out for the predictors Fixation Duration, Preceding Saccade Size and Current Saccade Size; see below).

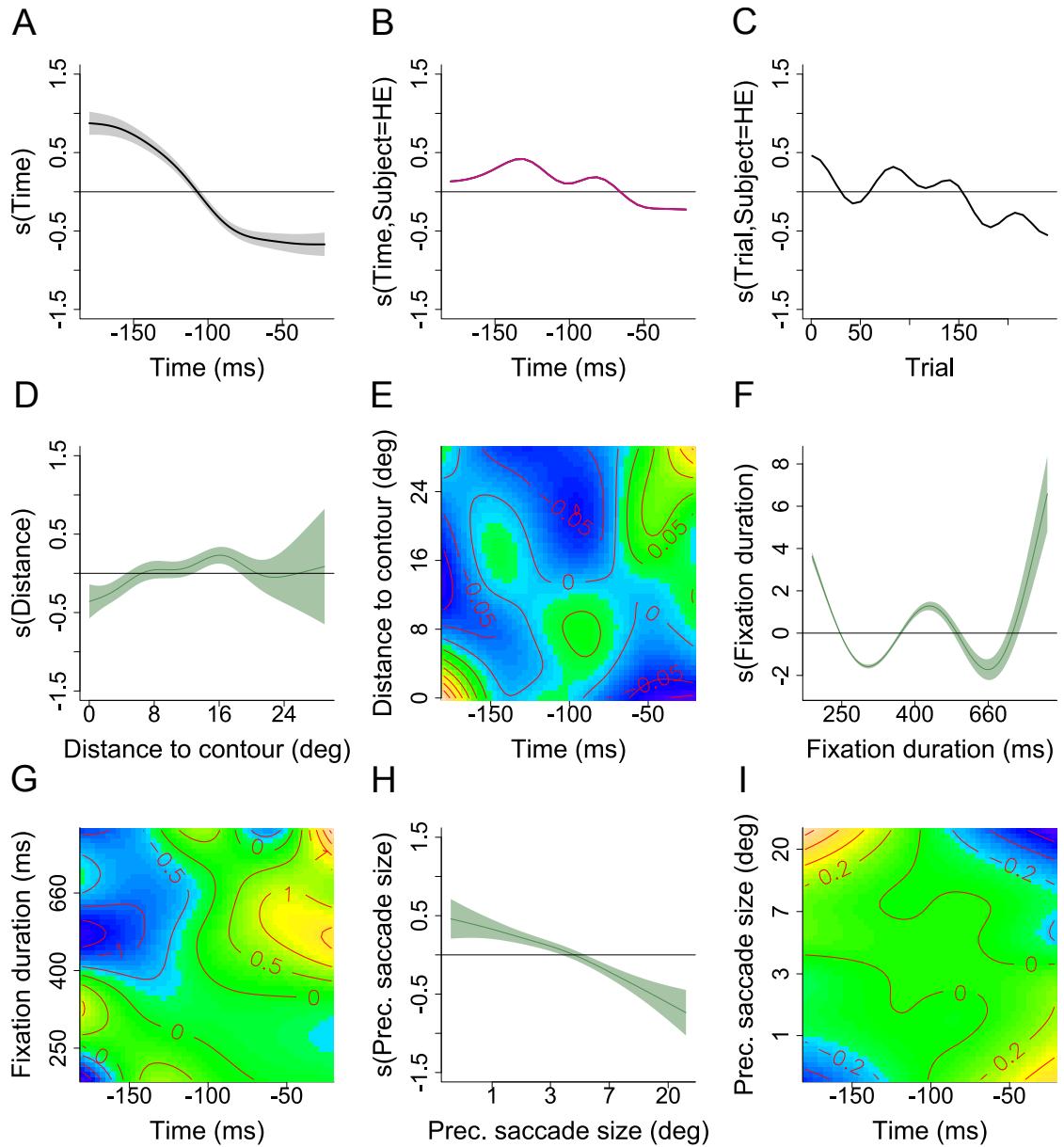


Figure 3.3. The GAMM components of a representative participant for the presaccadic EEG amplitude over the left occipital region for contour-present trials. Each graph represents the partial effect of a particular smooth term (for one numerical predictor) or tensor product (interaction between two numerical predictors). The first row shows (A) the partial effect of presaccadic Time and the random effects of (B) Time and (C) Trial for participants. The random intercept effect for Image is not depicted here. The second row depicts (D) the partial effect of Distance between the saccade landing position and the contour, (E) the partial interaction of Distance and presaccadic Time and (F) the partial effect of the covariate Fixation Duration. The last row shows (G) the partial interaction of Fixation Duration and presaccadic Time, (H) the partial effect of the covariate Preceding Saccade Size and (I) the partial interaction of Preceding Saccade Size and presaccadic Time.

3.3 Results

3.3.1 Response and eye movement behavior

Average hit rate across participants was 75% ($SD = 8.5\%$) and average false positive rate was 13% ($SD = 6.8\%$). Since we only analysed the trials in which a correct response was given, this resulted in the exclusion of 15% of the contour-present trials and 13% of the contour-absent trials.

Since our main variables of interest were Distance from the landing position of a saccade to the contour (for the contour-present trials) and Association Strength at the landing position of a saccade (for the contour-absent trials), we wanted to ensure that the size of this saccade does not systematically vary with these variables. Linear mixed analysis showed that Saccade Size could not be predicted by either the variable Distance in the contour-present trials ($\chi^2 = 3.35$, $p = 0.07$) or the variable Association Strength in the contour-absent trials ($\chi^2 = 0.72$, $p = 0.39$). The ROC integral value, indicating how well the model can predict participants' fixation locations, was 0.55 for the contour-present trials and 0.54 for the contour-absent trials. Both integral values were significantly different from the value that would be obtained by chance (i.e., 0.5, compared with a threshold of 0.53 for $p < 0.05$ based on bootstrapping simulations).

3.3.2 Results for contour-present trials

Random-effect variables Participants and Image

For the random-effect variable Participant, the GAMM revealed a significant nonlinear factor smooth for Time and Trial Number. The time course of EEG activity within the presaccadic epoch differed significantly between participants ($p < 0.001$). Trial-by-trial changes of EEG amplitude in the course of the experiment also significantly varied between participants ($p < 0.001$). In addition, we found a significant random intercept for the random-effect variable Image, revealing systematic differences in EEG amplitude

across images ($p < 0.001$; see Appendix A in section 3.8 for plots showing these structural variations for Participant and Image).

Distance between saccade landing position and contour

One of our predictors of interest was Distance between saccade landing position and contour. The partial effect of Distance was significant ($p < 0.001$). Inspection of p -values of smooth terms showed a significant partial effect of Distance for the left frontal, left parietal and both occipital regions ($p < 0.001$; Figure 3.4). For these regions, the GAMM model predicts EEG amplitude to increase with Distance. The most prominent effect was observed over the parieto-occipital regions. We therefore illustrate the following results with plots from the right occipital region, although the statistical analyses were always conducted on all eight ROIs.

The interaction between Distance and Time contributed significantly to the variance in EEG amplitude ($p < 0.001$). Examination of p -values for the tensor products revealed a significant partial interaction effect for both occipital regions ($p < 0.001$). As shown in Figure 3.5A, for large distances EEG amplitudes are lower at the start of the presaccadic epoch, while higher towards the end; for small distances the effects are opposite.

The linear mixed model analysis described in section 3.3.1 showed that the size of the current saccade (i.e., the saccade following the presaccadic interval) did not systematically vary with the distance of the saccade landing position to the contour. Nevertheless, we wanted to ensure that the size of this saccade could not account for the partial effect of Distance to the contour and its interaction with Time. To that end, a control GAMM analysis was conducted on the presaccadic EEG amplitude for the right occipital region, where the main GAMM analysis revealed the most prominent effect. In this analysis, the partial effect of Current Saccade Size and its interaction with Time were included. The partial effect of Current Saccade Size was significant ($p < 0.001$): EEG amplitudes increased with increasing saccade size. The partial interaction effect of Time and Current Saccade Size was also significant ($p < 0.001$): higher EEG amplitudes emerged over time for large saccade sizes, while the reversed effect occurred for small saccade sizes. The partial effect of our variable of interest, Distance, was still highly significant ($p < 0.001$),

as well as its partial interaction with Time ($p < 0.001$). Similar to the main analysis of all eight ROIs, higher EEG amplitudes occurred for larger distances to the contour. The partial interaction of Distance and Time was also similar to that observed in the main analysis.

Effects of eye movement parameters

We examined for the eight ROIs whether the eye movement covariates Fixation Duration and Preceding Saccade Size contributed to the variance in the EEG amplitude. Fixation

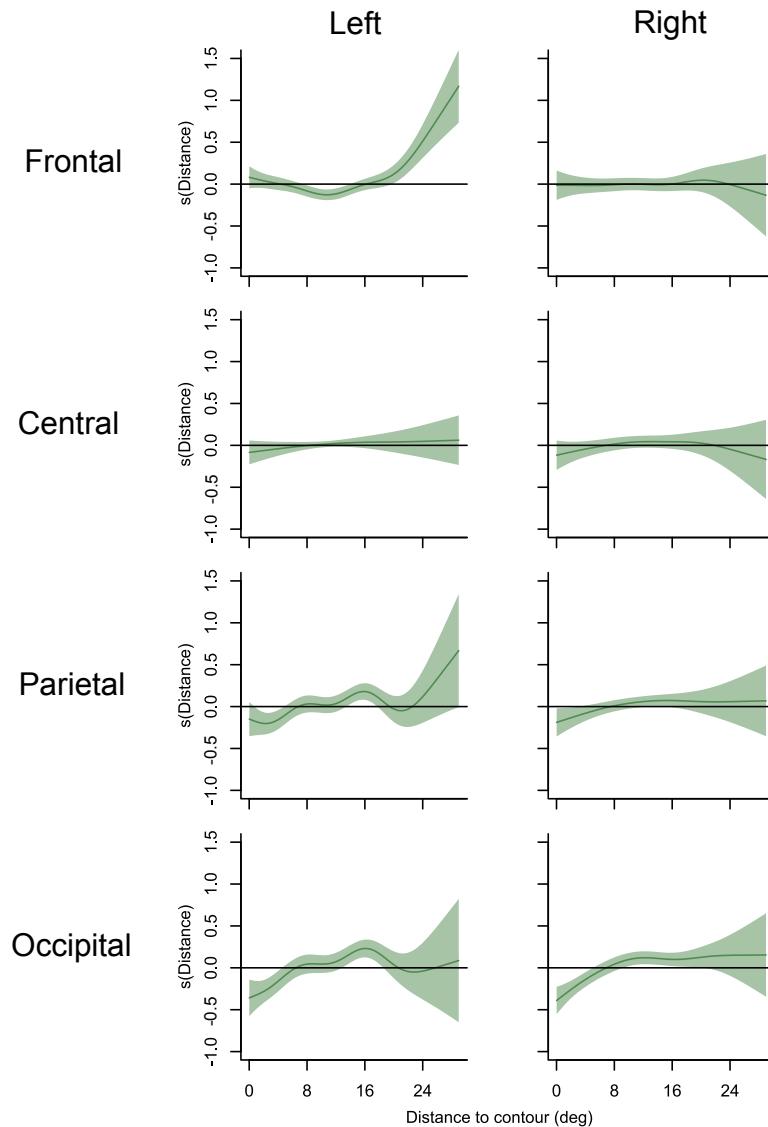


Figure 3.4. Contour-present trials. Partial effects of Distance between the saccade landing point and the contour on presaccadic EEG amplitude for the eight ROIs. Shaded regions depict two standard errors above and below the mean.

Duration was a highly significant predictor of presaccadic activity ($p < 0.001$ and $p < 0.001$ for each individual smooth term of each ROI). The EEG amplitude effects for this predictor showed a wave-like pattern, most prominent over frontal and occipital regions. For the frontal regions, higher EEG amplitudes occurred for fixation durations around 400 ms. For the occipital regions, higher EEG amplitudes emerged for fixations duration around 200 ms, 500 ms and 800 ms. Figure 3.5B depicts the partial effect of Fixation Duration for the right occipital region.

The partial interaction effect of Fixation Duration and Time on EEG amplitude was also

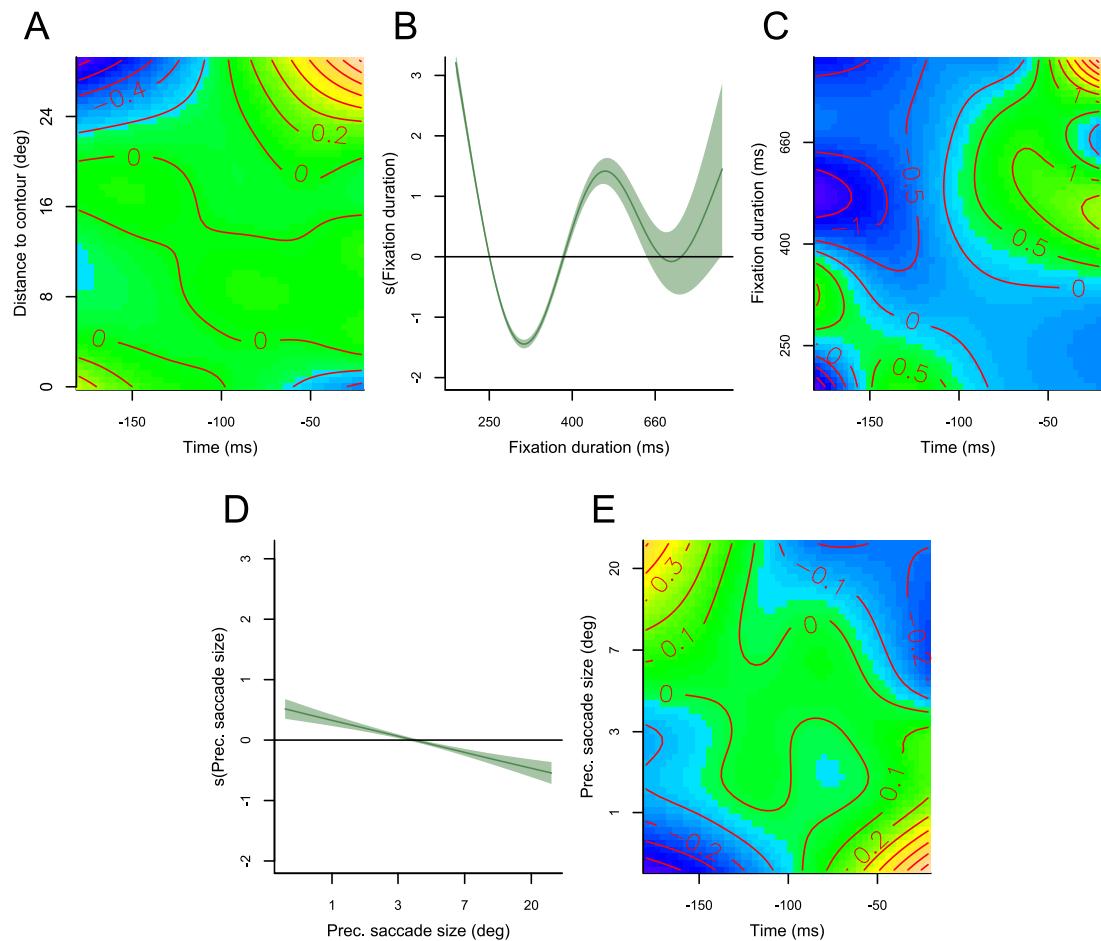


Figure 3.5. Contour-present trials. GAMM components for the right occipital ROI: (A) the partial interaction of presaccadic Time and Distance to the contour, (B) the partial effect of Fixation Duration, (C) the partial interaction of Time and Fixation Duration, (D) the partial effect of Preceding Saccade Size and (E) the partial interaction of Time and Preceding Saccade Size. In A, C, and E blue color denotes negative partial effects, whereas yellow and green colors indicate positive partial effects. Shaded regions in B and D denote two standard errors above and below the mean. Fixation Duration and Preceding Saccade Size are plotted on a logarithmic scale.

significant ($p < 0.001$ for each individual tensor product of each ROI). For all ROIs, the GAMM model predicts low EEG amplitudes at the beginning of the presaccadic epoch for fixations of medium and long duration, while high EEG amplitudes are predicted towards the end of the epoch. For fixation duration around 200 ms (the minimal allowed fixation duration in our analysis), low EEG amplitudes were present at the start of the presaccadic epoch over parietal and occipital regions. Figure 3.5C shows the partial interaction effect of Fixation Duration and Time for the right occipital region.

The partial effect of Preceding Saccade Size on presaccadic EEG amplitude was also significant ($p < 0.001$). Inspection of p -values of individual smooth terms showed that the effect was significant for left frontal, left central and occipital regions ($p < 0.001$). The GAMM model predicted increasing EEG amplitudes with increasing saccade size for the frontal regions, while the reversed direction of the effect was observed for the occipital regions. The partial effect of Preceding Saccade Size is shown in Figure 3.5D for the right occipital region.

The GAMM also revealed a significant partial interaction effect of Preceding Saccade Size and Time on EEG amplitude ($p < 0.001$). Inspection of p -values for the tensor products revealed a significant partial interaction effect for the frontal regions, left central and occipital regions ($p < 0.001$). For the occipital regions, higher EEG amplitudes occurred at the start of the presaccadic interval while lower EEG amplitudes occurred towards the end of the epoch for large saccade sizes. The opposite effect occurred for small saccade sizes. For the frontal regions, the opposite direction of the effects for large saccade sizes was observed. Figure 3.5E depicts the partial interaction effect of Preceding Saccade Size and Time on EEG amplitude for the right occipital region.

3.3.3 Results for contour-absent trials

Random-effect variables Participants and Image

The random effects for the contour-absent trials yielded results similar to the contour-present trials. For Participant, a significant nonlinear factor smooth was found for Time and Trial Number. The presaccadic time and trial course of EEG activity differed signifi-

cantly between participants (both $p < 0.001$). In addition, we found a significant random intercept for the random-effect variable Image, indicating systematic variation in EEG amplitude across images ($p < 0.001$; see Appendix B in section 3.9 for plots showing these structural variations for Participant and Image).

Association strength at the saccade landing position

The GAMM showed a significant partial effect of the predictor Association Strength ($p < 0.001$). Inspection of p -values showed that the partial effect of Association Strength was highly significant for all brain regions ($p < 0.001$). The partial effects of Association Strength on presaccadic activity for each ROI are shown in Figure 3.6. Higher EEG amplitudes characterize low association strengths for the occipital regions. This observation corresponds to the higher EEG amplitudes which occur with larger distances to the contour in the contour-present trials.

The partial interaction effect of Association Strength and Time on the presaccadic EEG amplitude was significant according to a Chi-Square test of maximum-likelihood scores. Examination of p -values for the tensor products, however, indicated no significant partial interaction effects in any of the ROIs. Due to the inconsistent results between the two methods, we consider these effects insignificant. Figure 3.7A shows the partial interaction effect of presaccadic Time and Association Strength for the right occipital region.

Similar to our analysis of presaccadic EEG amplitude in the contour-present trials, we wanted to ensure that the size of the saccade following the presaccadic interval (i.e., the Current Saccade Size) could not account for the partial effect of Association Strength and its interaction with Time. A control GAMM analysis was therefore conducted on the presaccadic EEG amplitude for the right occipital region, in which the partial effect of Current Saccade Size and its interaction with Time were included. The partial effect of Current Saccade Size was significant ($p < 0.001$): presaccadic EEG amplitudes increased with increasing saccade size. The partial interaction effect of Time and Current Saccade Size was also significant ($p < 0.001$). However, no clear pattern arose as the wiggliness of the estimated smoothing function was high. Importantly, the partial effect of Association Strength was still highly significant ($p < 0.001$), as well as its partial interaction with

Time ($p < 0.001$). Similar to the main analysis of all eight ROIs, higher EEG amplitudes occurred again for low association strengths. The partial interaction of Association Strength and Time was also similar to that observed in the main analysis.

Effects of eye movement parameters

Similarly to the contour-present trials, GAMM revealed a significant partial effect of Fixation Duration on the amplitude of presaccadic activity ($p < 0.001$ and $p < 0.001$ for each individual smooth term of each ROI). The GAMM model predicted a wave-like

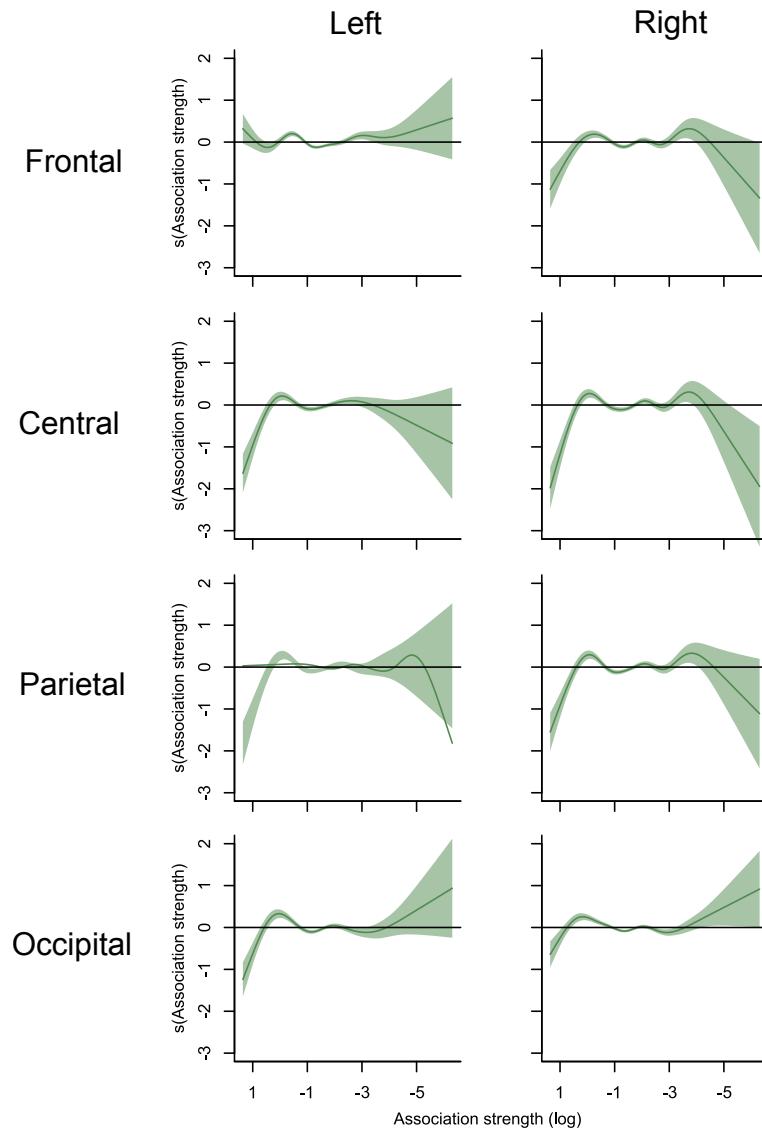


Figure 3.6. Contour-absent trials. Partial effects of Association Strength of the saccade landing point on presaccadic EEG amplitude for each ROI. Shaded regions depict two standard errors above and below the mean. Association Strength is plotted on a logarithmic scale.

pattern across fixation durations with smallest wiggleness for the central regions. For the parietal and occipital regions, EEG amplitudes peaked for fixation durations of about 200, 550 and 800 ms. On the contrary, for the frontal regions, EEG amplitudes peaked only once, around 400 ms. For the central regions, EEG amplitudes slightly increased with increasing fixation durations. Figure 3.7B depicts the partial effect of Fixation Duration on EEG amplitude for the right occipital region.

A partial interaction between Fixation Duration and Time was also significant ($p < 0.001$ and $p < 0.001$ for each individual tensor product of each ROI). For all regions although

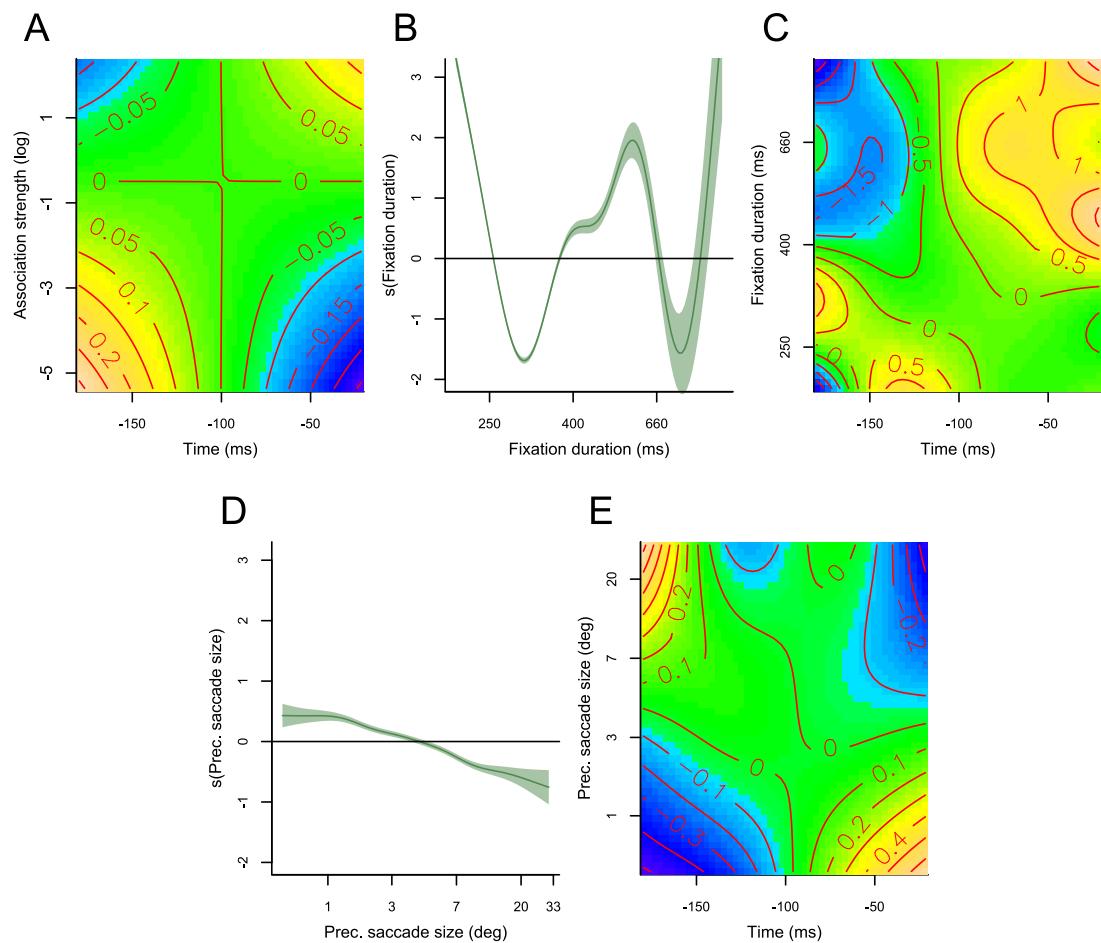


Figure 3.7. Contour-absent trials. GAMM components for the right occipital ROI: (A) partial interaction of presaccadic Time and Association Strength, (B) the partial effect of Fixation Duration, (C) the partial interaction effect of Time and Fixation Duration, (D) the partial effect of Preceding Saccade Size and (E) the partial interaction effect of Time and Preceding Saccade Size. In A, C, E the blue color denotes negative partial effects, whereas yellow and green colors indicate positive partial effects. Shaded regions in B and D denote two standard errors above and below the mean. Fixation Duration and Preceding Saccade Size are plotted on a logarithmic scale.

least visible over the frontal ones, lower EEG amplitudes occurred at the beginning of the presaccadic epoch for long fixation durations, while higher EEG amplitudes appear towards the end of the epoch. The GAMM predicted lower EEG amplitudes at the beginning of the epoch for fixation durations around 200 ms. These effects are similar to those observed in the contour-present trials. The partial interaction effect of presaccadic Time and Fixation Duration on EEG amplitude for the right occipital region is shown in Figure 3.7C.

The partial effect of Preceding Saccade Size on EEG amplitude was significant ($p < 0.001$). Significant effects were found for all regions ($p < 0.001$) except for the left central one ($p = 0.53$). For the parietal and occipital regions, higher EEG amplitudes emerged for small saccade sizes, while lower EEG amplitudes occurred for large saccade sizes. This negative relationship between the amplitude effects and saccade size was similar to that observed in contour-present trials. For the frontal and central ROIs, higher EEG amplitudes occurred for increasing Preceding Saccade Size with slightly smaller effects for medium saccade sizes than for large or small saccade sizes. Figure 3.7D depicts the partial effect of Preceding Saccade Size on EEG amplitude for the right occipital region.

GAMM also revealed a significant partial interaction effect on EEG amplitude of Preceding Saccade Size and Time ($p < 0.001$). Inspection of p -values for the tensor products revealed significant effects for the frontal, parietal and occipital regions ($p < 0.001$). For the occipital and parietal regions, EEG amplitudes decreased with presaccadic time for large saccade sizes and increased with time for small saccade sizes. The direction of effects observed for large saccade sizes was opposite for the frontal channels. The partial interaction effect of presaccadic Time and Preceding Saccade Size on EEG amplitude for the right occipital region is shown in Figure 3.7E.

3.3.4 Variance explained by GAMM

It should be noted that, although the effects reported are highly significant, the proportion variance explained by the model is low: for the GAMM analysis of all eight ROIs, the proportion variance explained is only 2.4% for the contour-present trials and 2.9% for the

contour-absent trials. For the GAMM analysis of the right occipital region, the proportion variance explained is 3% for the contour-present trials and 4.5% for the contour-absent trials. These values are comparable with the values reported in studies which analyse single-epoch EEG data (Dambacher, Kliegl, Hofmann, & Jacobs, 2006; Meulman et al., 2015; Payne, Lee, & Federmeier, 2015; Tremblay & Newman, 2015). The explained variance reported in EEG studies which *average* over multiple trials to obtain an ERP signal is typically higher (although many EEG studies do not even report the variance explained or any other indication of effect size). This is due to the fact that averaging over multiple trials inevitably leads to the removal of individual variability across EEG epochs, trials and subjects and therefore leads to an overestimation of the amount of explained variability. Importantly, this averaging does not necessarily remove variability merely due to noise, but it may also mask systematic variability relevant to the research questions (Meulman et al., 2015; Payne et al., 2015). In contrast, the GAMM approach considers all these potential sources of variability. Consequently, the proportion variance explained will be much lower, but it will provide a more realistic account of the true variability that is present in the EEG signal.

3.4 Discussion

In this study, we examined the attentional brain processes involved in the selection of saccade targets during free-viewing of a display. Participants searched for a contour of collinear Gabor elements embedded in a background of randomly oriented but otherwise similar distractors. Saccade-related EEG and eye movements were simultaneously recorded. We found that, when a contour was present, presaccadic EEG activity over parieto-occipital brain areas predicted the proximity of the upcoming fixation location to the contour. When a contour was absent, the presaccadic amplitude predicted the association strength at the upcoming fixation location.

Presaccadic EEG activity was found to be larger when an upcoming fixation location was associated with low saliency. As mentioned earlier, presaccadic EEG activity seems to reflect the level of attention required to select the upcoming saccade target (i.e., the next

fixation location) and to ignore other locations (Ptak et al., 2011). The presaccadic activity thus reflects an attentional shift to the next fixation target, away from other potential targets (Gutteling et al., 2010; Krebs et al., 2012; Wauschkuhn et al., 1998). Therefore, our findings suggest that more attention is being deployed for the selection of less salient objects. It is known that salient regions trigger reflexive saccades automatically and that they involve bottom-up visual processing (Itti & Koch, 2001; Parkhurst et al., 2002). In other words, salient locations attract saccades with ease, while non-salient locations might require more “effort”. This is either to compensate for lack of attraction or in order to overcome the competition with more attractive candidates. Saccades away from salient areas may result from top-down control processes winning the competition with bottom-up saliency (Cosman & Vecera, 2013; Kristjánsson, 2007). Such saccades probably serve an exploratory function and might indicate a serial search strategy (Cosman & Vecera, 2013).

The validity of our conclusions is supported by the statistical power of single-trial analysis, which is still rare in modern EEG research. This analysis was made possible by using GAMM. Another property of GAMM, which makes it particularly valuable for EEG research, is its ability to account for nonlinearity of time series data. In addition, by including random-effect variables, the model enables us to appropriately deal with systematic variation between participants and images in EEG amplitude (see De Cat, Klepousniotou, & Baayen, 2015; Kryuchkova, Tucker, Wurm, & Baayen, 2012; Meulman et al., 2015 for other ERP research that applied GAMM).

The main reason for the application of GAMM in our study was related to the fundamental problem of EEG-eye movement co-registration in free-viewing behavior. The analysis of EEG in free-viewing allows to evaluate eye movement behavior during contour integration under more ecologically valid conditions than in the traditional stimulus-response paradigm. However, it produces a serious complication for the simultaneous EEG-eye movement analysis. In such analysis, the EEG signal is segmented relatively to saccades and fixations. Due to the sequential nature of eye movements, the saccadic effects on the current EEG segment overlap with the effects from the preceding saccades. Previous attempts to solve this problem had different drawbacks. For instance, Dimigen et al.

(2011) examined brain potentials time-locked to the fixation onset under natural reading conditions and applied a Generalized Linear Mixed Model (GLMM) approach (of which the GAMM approach is an extension; see Methods section) to control for the size of saccades distorting the EEG waveform during the following fixation. However, this technique could not effectively deal with the nonlinear distortions resulting from overlap of EEG responses to previous saccades. Such distortions are nonlinear since each overlapping potential constitutes a complex waveform. As we have discussed in the introduction, the distortions are predominantly caused by lambda responses, which depend on fixation duration and saccade size (see also below).

Another approach to account for the differential overlapping effect of previous and subsequent saccades involves the use of a GLMM-based deconvolution method (Dandekar et al., 2012). In general, deconvolution techniques attempt to separate overlapping potentials which occur in the EEG waveform in the course of fast-paced stimulation (Jewett et al., 2004; Wang, Özdamar, Bohórquez, Shen, & Cheour, 2006). These techniques require a template waveform which is supposed to be identical for each successive stimulation. Dandekar et al. (2012) modelled the EEG signal separately for different time samples of the saccadic main sequence. The EEG activity at a given time relative to a saccade onset is then estimated as the linear superposition of the responses estimated in preceding and following time samples. To account for different EEG responses to different saccade size, deconvolution was done separately for five ranges of saccade size. In such a way, the overlapping influences of saccade size were efficiently minimized. However, this study did not consider other eye movement characteristics which also contribute to overlapping effects, such as fixation duration. In addition, the deconvolution method heavily relies on knowledge about a template of the non-overlapping EEG signal related to a fixation or saccade (Dimigen et al., 2011).

To remove EEG responses evoked by eye movements, Dias et al. (2013) used a subspace subtraction technique, which is similar to ICA, except that it does not identify components “blindly” but pre-specifies certain properties of these components. In their study, the components associated with saccadic spike potential, and vertical and horizontal eye movements were identified and removed from the EEG signal. However, this method

is also subjected to the problem of differential overlap of EEG responses to subsequent saccades and hence cannot reliably “single out” the components.

The GAMM approach allowed us to disentangle the effects of eye movement covariates on presaccadic EEG activity from our main variables of interest. As expected, we found that fixation duration produced systematic distortions in the presaccadic amplitude. The effect of fixation duration reflects the effect of the lambda response occurring approximately 100 ms after the offset of the previous saccade. The latency of the lambda response and its specific overlap with the baseline and presaccadic interval changes with fixation duration, resulting in a wave-like effect of fixation duration on the presaccadic EEG amplitude (Dimigen, 2014; Nikolaev et al., 2016). The preceding saccade size was also observed to significantly contribute to the variance in the presaccadic amplitude. We found that presaccadic EEG activity decreased with increasing saccade size. This is again to be expected since it is known that the lambda response and saccadic spike potential (i.e., the potential reflecting eye muscle activity at the saccade onset) increase with increasing saccade size (Thickbroom et al., 1991; Yagi, 1979). The effect of preceding saccade size is then presumably due to the fact that our baseline interval overlaps with either part of the lambda response or saccadic spike potential. Larger saccadic spike potentials or lambda responses will result in larger baseline corrections. Consequently, lower presaccadic activity is observed relative to this baseline interval (see Dimigen, 2014; Nikolaev et al., 2016 for a more detailed overview of the possible effects of these eye movement parameters).

It should be noted that we also found an effect of the following saccade size on the presaccadic EEG amplitude, which was unexpected based on previous work (Nikolaev et al., 2016). Dependence between the presaccadic amplitude and the size of the following saccade may reflect saccade planning and oculomotor preparation for saccade execution. Thus, the presaccadic EEG amplitude may be affected by a mixture of factors, such as attention and oculomotor preparation. A second point is that the effects due to overlap in EEG responses involve not only the EEG response to the immediately preceding saccade, but also to earlier eye movements. We chose not to include the characteristics of these earlier eye movements in the model because their effect was considered relatively weak.

Moreover, they increase the complexity of the model considerably which would have rendered model selection infeasible due to the required computing time.

In sum, the goal of this study was to test the hypothesis that presaccadic EEG amplitude can predict visual saliency at the subsequent fixation location during a contour integration task. Our findings were based on an elaborate GAMM analysis, which is a relatively novel method to analyze ERP data. We showed that GAMM is highly suitable for analyzing the presaccadic EEG amplitude by offering a solution to the problem of differential overlap due to subsequent saccades. Further research will be necessary to further prove the usefulness of GAMM in analyzing saccade-related and fixation-related EEG potentials.

3.5 Appendix A: Grand average and random effects for contour-present trials

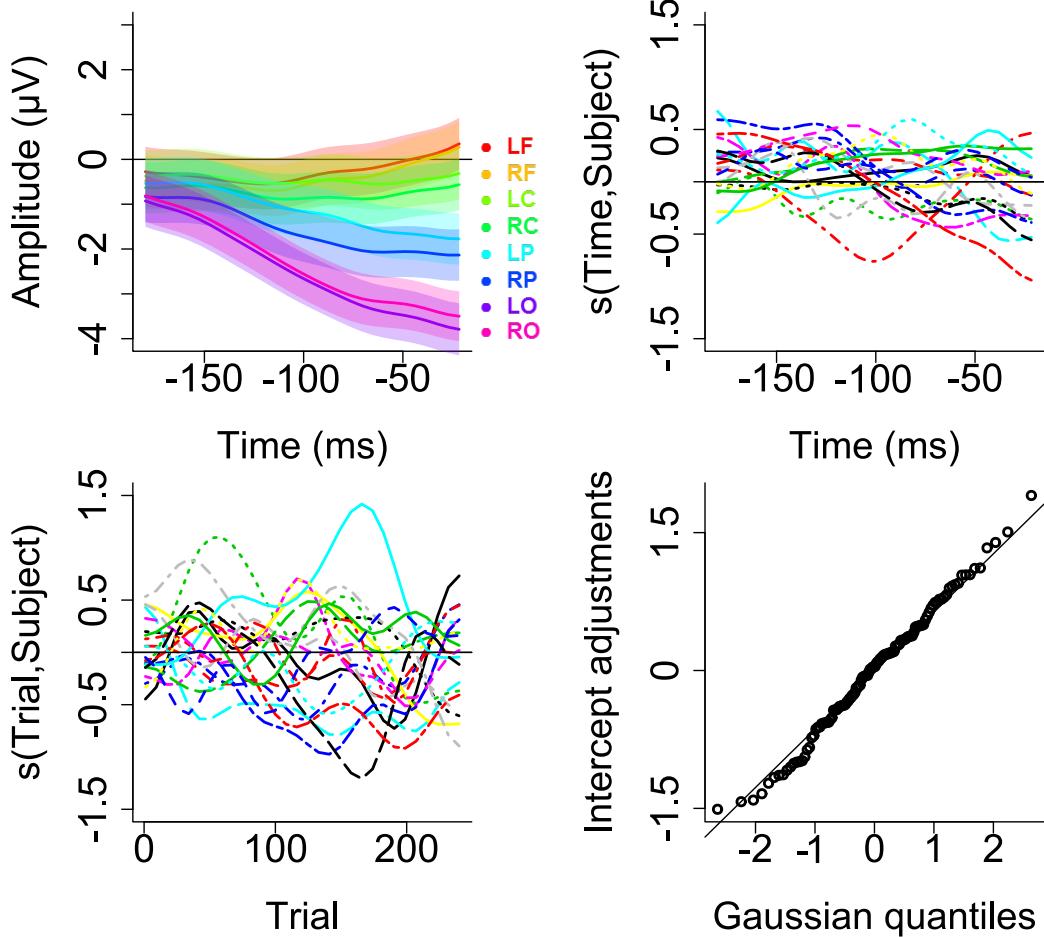


Figure 3.8. Contour-present trials. (A) Average amplitude for each ROI as function of presaccadic Time. Each colored line represents a ROI: LF = Left Frontal area, RF = Right Frontal area, LC = Left Central area, RC = Right Central area, LP = Left Parietal area, RP = Right Parietal area, LO = Left Occipital area, RO = Right Occipital area. (B) Random effect for Participant as a function of presaccadic Time. Each colored line depicts a participant. (C) Random effect for Participant as a function of Trial Number. Each colored line depicts a participant. (D) The adjustments to the baseline EEG amplitude as a function of Image.

3.6 Appendix B: Grand average and random effects for contour-absent trials

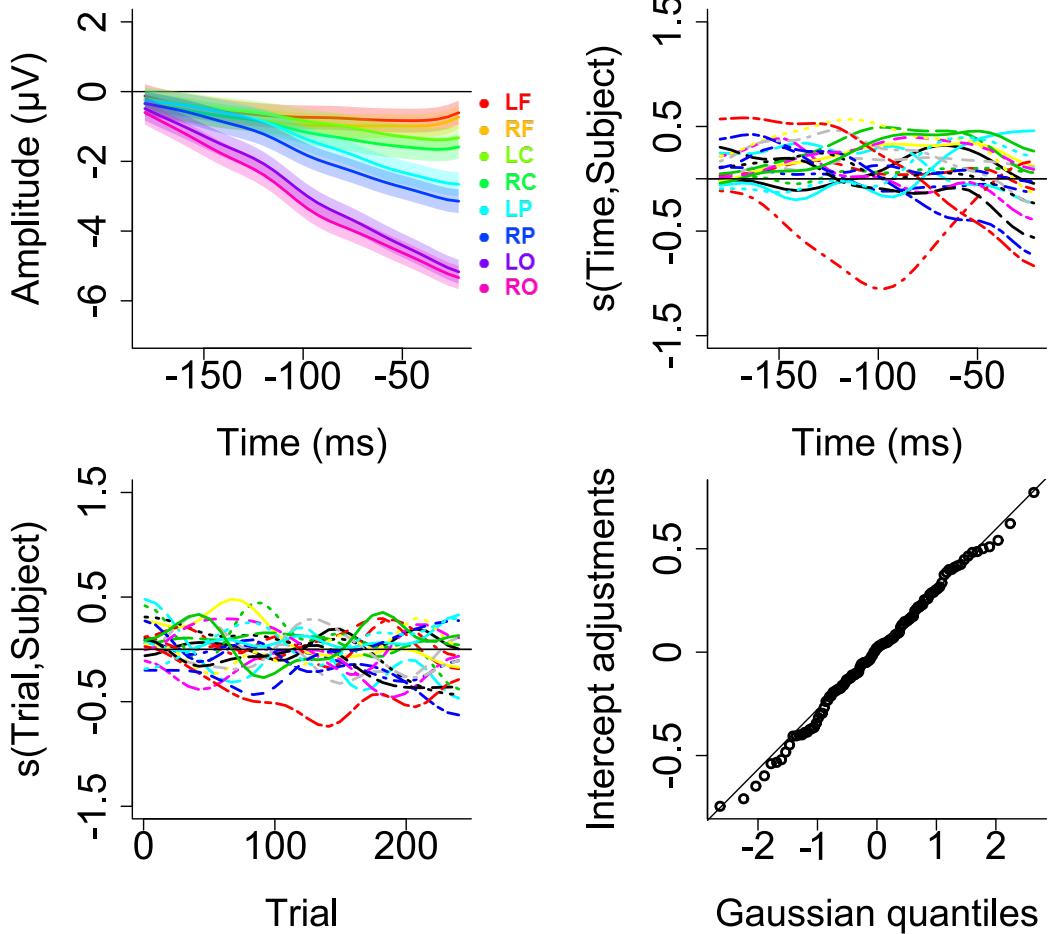


Figure 3.9. Contour-absent trials. (A) Average amplitude for each ROI as function of presaccadic Time. Each colored line represents a ROI: LF = Left Frontal area, RF = Right Frontal area, LC = Left Central area, RC = Right Central area, LP = Left Parietal area, RP = Right Parietal area, LO = Left Occipital area, RO = Right Occipital area. (B) Random effect for Participant as a function of presaccadic Time. Each colored line depicts a participant. (C) Random effect for Participant as a function of Trial Number. Each colored line depicts a participant. (D) The adjustments to the baseline EEG amplitude as a function of Image.

References

- Alexander, D. M., Jurica, P., Trengove, C., Nikolaev, A. R., Gepshtein, S., Zvyagintsev, M., ... van Leeuwen, C. (2013). Traveling waves and trial averaging: The nature of single-trial and averaged brain responses in large-scale cortical signals. *Neuroimage*, 73, 95–112.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1–48.
- Brouwer, A.-M., Reuderink, B., Vincent, J., van Gerven, M. A., & van Erp, J. B. (2013). Distinguishing between target and nontarget fixations in a visual search task using fixation-related potentials. *Journal of Vision*, 13(3), 1–10.
- Cohen, M. X. (2014). *Analyzing neural time series data: theory and practice*. Cambridge, Massachusetts MIT Press.
- Cosman, J. D., & Vecera, S. P. (2013). Context-dependent control over attentional capture. *Journal of Experimental Psychology: Human Perception and Performance*, 39(3), 836–848.
- Dambacher, M., Kliegl, R., Hofmann, M., & Jacobs, A. M. (2006). Frequency and predictability effects on event-related potentials during reading. *Brain Research*, 1084(1), 89–103.
- Dandekar, S., Privitera, C., Carney, T., & Klein, S. A. (2012). Neural saccadic response estimation during natural viewing. *Journal of Neurophysiology*, 107(6), 1776–1790.
- De Cat, C., Klepousniotou, E., & Baayen, R. H. (2015). Representational deficit or processing effect? An electrophysiological study of noun-noun compound processing by very advanced L2 speakers of English. *Frontiers in Psychology*, 6(77), 1–17.
- Demeyer, M., & Machilsen, B. (2012). The construction of perceptual grouping displays using GERT. *Behavior Research Methods*, 44(2), 439–446.
- Devillez, H., Guyader, N., & Guérin-Dugué, A. (2015). An eye fixation-related potentials analysis of the P300 potential for fixations onto a target object when exploring natural scenes. *Journal of Vision*, 15(13), 1–31.
- Dias, J. C., Sajda, P., Dmochowski, J. P., & Parra, L. C. (2013). EEG precursors of detected and missed targets during free-viewing search. *Journal of Vision*, 13(13), 1–19.
- Dimigen, O. (2014). *Co-registration of eye movements and EEG during active vision* (Unpublished doctoral dissertation). Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät II.
- Dimigen, O., Sommer, W., Hohlfeld, A., Jacobs, A. M., & Kliegl, R. (2011). Coregistration of eye movements and EEG in natural reading: analyses and review. *Journal of Experimental Psychology: General*, 140(4), 552–572.
- Ernst, U. A., Mandon, S., Schinkel-Bielefeld, N., Neitzel, S. D., Kreiter, A. K., & Pawelzik, K. R. (2012). Optimality of human contour integration. *PLoS Computational Biology*, 8(5), e1002520.
- Field, D. J., Hayes, A., & Hess, R. F. (1993). Contour integration by the human visual system: evidence for a local “association field”. *Vision Research*, 33(2), 173–193.
- Foulsham, T., & Underwood, G. (2008). What can saliency models predict about eye movements? Spatial and sequential aspects of fixations during encoding and recog-

- nition. *Journal of Vision*, 8(2), 1–17.
- Gratton, G., Coles, M. G., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology*, 55(4), 468–484.
- Gutteling, T. P., van Ettinger-Veenstra, H. M., Kenemans, J. L., & Neggers, S. F. (2010). Lateralized frontal eye field activity precedes occipital activity shortly before saccades: evidence for cortico-cortical feedback as a mechanism underlying covert attention shifts. *Journal of Cognitive Neuroscience*, 22(9), 1931–1943.
- Hastie, T. J., & Tibshirani, R. J. (1990). *Generalized additive models* (Vol. 43). CRC Press.
- Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences*, 7(11), 498–504.
- Itti, L., & Koch, C. (2001). Computational modeling of visual attention. *Nature Reviews Neuroscience*, 2(3), 194–203.
- Jewett, D. L., Caplovitz, G., Baird, B., Trumppis, M., Olson, M. P., & Larson-Prior, L. J. (2004). The use of QSD (q-sequence deconvolution) to recover superposed, transient evoked-responses. *Clinical neurophysiology*, 115(12), 2754–2775.
- Jung, T.-P., Makeig, S., Humphries, C., Lee, T.-W., McKeown, M. J., Iragui, V., & Sejnowski, T. J. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*, 37(2), 163–178.
- Jung, T.-P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., & Sejnowski, T. J. (2000). Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. *Clinical Neurophysiology*, 111(10), 1745–1758.
- Kamienkowski, J. E., Ison, M. J., Quiroga, R. Q., & Sigman, M. (2012). Fixation-related potentials in visual search: A combined EEG and eye tracking study. *Journal of Vision*, 12(7), 1–20.
- Kaunitz, L. N., Kamienkowski, J. E., Varatharajah, A., Sigman, M., Quiroga, R. Q., & Ison, M. J. (2014). Looking for a face in the crowd: Fixation-related potentials in an eye-movement visual search task. *Neuroimage*, 89, 297–305.
- Kelly, S. P., Foxe, J. J., Newman, G., & Edelman, J. A. (2010). Prepare for conflict: EEG correlates of the anticipation of target competition during overt and covert shifts of visual attention. *European Journal of Neuroscience*, 31(9), 1690–1700.
- Keren, A. S., Yuval-Greenberg, S., & Deouell, L. Y. (2010). Saccadic spike potentials in gamma-band EEG: characterization, detection and suppression. *Neuroimage*, 49(3), 2248–2263.
- Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiology*, 4(4), 219–227.
- Körner, C., Braunstein, V., Stangl, M., Schlögl, A., Neuper, C., & Ischebeck, A. (2014). Sequential effects in continued visual search: Using fixation-related potentials to compare distractor processing before and after target detection. *Psychophysiology*, 51(4), 385–395.
- Kotowicz, A., Rutishauser, U., & Koch, C. (2010). Time course of target recognition in visual search. *Frontiers in Human Neuroscience*, 4(31), 1–11.
- Krebs, R. M., Boehler, C. N., Zhang, H. H., Schoenfeld, M. A., & Woldorff, M. G. (2012). Electrophysiological recordings in humans reveal reduced location-specific attentional-shift activity prior to recentering saccades. *Journal of Neurophysiology*,

107(5), 1393–1402.

- Kristjánsson, A. (2007). Saccade landing point selection and the competition account of pro-and antisaccade generation: The involvement of visual attention—A review. *Scandinavian Journal of Psychology*, 48(2), 97–113.
- Kryuchkova, T., Tucker, B. V., Wurm, L. H., & Baayen, R. H. (2012). Danger and usefulness are detected early in auditory lexical processing: Evidence from electroencephalography. *Brain and Language*, 122(2), 81–91.
- Lins, O. G., Picton, T. W., Berg, P., & Scherg, M. (1993). Ocular artifacts in recording EEGs and event-related potentials II: Source dipoles and source components. *Brain Topography*, 6(1), 65–78.
- Makeig, S., Bell, A. J., Jung, T.-P., & Sejnowski, T. J. (1996). Independent Component Analysis of Electroencephalographic Data. In D. Touretzky, M. Mozer, & M. Hasselmo (Eds.), *Advances in neural information processing systems* (Vol. 8, pp. 145–151). Cambridge, MA: The MIT Press.
- Mazaheri, A., DiQuattro, N. E., Bengson, J., & Geng, J. J. (2011). Pre-stimulus activity predicts the winner of top-down vs. bottom-up attentional selection. *PLoS One*, 6(2), e16243.
- Meulman, N., Wieling, M., Sprenger, S. A., Stowe, L. A., & Schmid, M. S. (2015). Age effects in L2 grammar processing as revealed by ERPs and how (not) to study them. *PLoS One*, 10(12), e0143328.
- Nikolaev, A. R., Jurica, P., Nakatani, C., Plomp, G., & van Leeuwen, C. (2013). Visual encoding and fixation target selection in free viewing: presaccadic brain potentials. *Frontiers in Systems Neuroscience*, 7(26), 1–12.
- Nikolaev, A. R., Meghanathan, R. N., & van Leeuwen, C. (2016). Combining EEG and eye movement recording in free viewing: Pitfalls and possibilities. *Brain and Cognition*, 107, 55–83.
- Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research*, 42(1), 107–123.
- Payne, B. R., Lee, C.-L., & Federmeier, K. D. (2015). Revisiting the incremental effects of context on word processing: Evidence from single-word event-related brain potentials. *Psychophysiology*, 52(11), 1456–1469.
- Ptak, R., Camen, C., Morand, S., & Schnider, A. (2011). Early event-related cortical activity originating in the frontal eye fields and inferior parietal lobe predicts the occurrence of correct and error saccades. *Human Brain Mapping*, 32(3), 358–369.
- Rothkopf, C. A., Ballard, D. H., & Hayhoe, M. M. (2007). Task and context determine where you look. *Journal of Vision*, 7(14), 1–20.
- Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: Reinterpreting salience. *Journal of Vision*, 11(5), 1–23.
- Thickbroom, G., Knezevic, W., Carroll, W., & Mastaglia, F. (1991). Saccade onset and offset lambda waves: relation to pattern movement visually evoked potentials. *Brain Research*, 551(1), 150–156.
- Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. *Psychological Review*, 113(4), 766–786.
- Tremblay, A., & Newman, A. J. (2015). Modeling nonlinear relationships in ERP data using mixed-effects regression with R examples. *Psychophysiology*, 52(1), 124–139.

- Van Humbeeck, N., Schmitt, N., Hermens, F., Wagemans, J., & Ernst, U. A. (2013). The role of eye movements in a contour detection task. *Journal of Vision*, 13(14), 1–19.
- van Rij, J., Wieling, M., Baayen, R. H., & van Rijn, H. (2015). *itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs*. (R package version 1.0.1)
- Vossen, H., Van Breukelen, G., Hermens, H., Van Os, J., & Lousberg, R. (2011). More potential in statistical analyses of event-related potentials: a mixed regression approach. *International journal of methods in psychiatric research*, 20(3), e56–e68.
- Wang, T., Özdamar, Ö., Bohórquez, J., Shen, Q., & Cheour, M. (2006). Wiener filter deconvolution of overlapping evoked potentials. *Journal of Neuroscience Methods*, 158(2), 260–270.
- Watt, R., Ledgeway, T., & Dakin, S. C. (2008). Families of models for Gabor paths demonstrate the importance of spatial adjacency. *Journal of Vision*, 8(7), 1–19.
- Wauschkuhn, B., Verleger, R., Wascher, E., Klostermann, W., Burk, M., Heide, W., & Kömpf, D. (1998). Lateralized human cortical activity for shifting visuospatial attention and initiating saccades. *Journal of Neurophysiology*, 80(6), 2900–2910.
- Wenzel, M. A., Golenia, J.-E., & Blankertz, B. (2016). Classification of eye fixation related potentials for variable stimulus saliency. *Frontiers in Neuroscience*, 10(23), 1–14.
- Wolfe, J. M., Butcher, S. J., Lee, C., & Hyle, M. (2003). Changing your mind: on the contributions of top-down and bottom-up guidance in visual search for feature singletons. *Journal of Experimental Psychology: Human Perception and Performance*, 29(2), 483–502.
- Wood, S. N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(1), 95–114.
- Wood, S. N. (2006). *Generalized additive models: an introduction with R*. CRC press.
- Wood, S. N. (2015). *mgcv: Mixed GAM Computation Vehicle with GCV/AIC/REML Smoothness Estimation*. (R package version 1.8.6)
- Yagi, A. (1979). Saccade size and lambda complex in man. *Physiological Psychology*, 7(4), 370–376.

Chapter 4

Apparent motion suppresses responses in V1: a population code model

This chapter has been submitted for publication as:

Van Humbeeck, N., Putzeys, T. & Wagemans, J. (2016). *Apparent motion suppresses responses in V1: a population code model*. Manuscript under revision.

4.1 Introduction

Apparent motion (AM) is a type of illusory motion that can be perceived when two stimuli are presented alternately at two different locations (Wertheimer, 1912). Under optimal spatial and temporal stimulus conditions, observers can perceive a single stimulus moving continuously along a path between the two locations. Some studies found neurons in primary visual cortex (V1) to respond during AM as if the stimulus was physically present at intermediate locations along the AM path (Larsen, Madsen, Ellegaard Lund, & Bundesen, 2006; Muckli, Kohler, Kriegeskorte, & Singer, 2005). It has been claimed that humans perceive AM because of these V1 responses, indicating that AM has an early cortical locus (Muckli et al., 2005). AM-related V1 responses may be the result of feedback from higher visual areas specialized in motion, such as hMT/V5+ (Ahmed et al., 2008; Vetter, Grosbras, & Muckli, 2015; Wibral, Bledowski, Kohler, Singer, & Muckli, 2009). These areas have large receptive fields, which allows for the representation of long-range AM. However, evidence for the neural mechanism underlying AM remains inconclusive. Liu, Slotnick, and Yantis (2004) failed to find AM-related activity in primary visual cortex during the percept of moving concentric rings. These authors did find increased responses in motion processing areas, suggesting that AM has a late cortical locus. Other studies report a similar lack of activation in V1 (Muckli et al., 2002; Goebel, Khorram-Sefat, Muckli, Hacker, & Singer, 1998; Mikami, Newsome, & Wurtz, 1986).

Behavioral studies have supported the hypothesis of AM-related activation in V1 by reporting impaired perception of stimuli presented along the path of AM (Hidaka, Nagai, Sekuler, Bennett, & Gyoba, 2011; Yantis & Nakama, 1998). For instance, Hidaka et al. (2011) induced an AM percept using Gabor gratings of a specific orientation. They observed that detectability of a target grating was impaired along the AM path, but only when the orientation of the target matched the orientation of the AM-inducing gratings. This AM masking has been explained by perceptual filling-in at the level of V1. The presentation of the AM inducers evokes responses in a subset of V1 cells, tuned to locations along the AM path and to the visual properties of these inducers. These evoked responses interfere with the responses to actual stimuli along the AM path, thus

impairing the perception of these stimuli.

However, AM masking does not necessarily imply V1 activation. An alternative view on AM masking can be provided by a predictive coding account of visual processing. Predictive coding emphasizes the notion of a “predictive brain” generating predictions about incoming information based on the surrounding context (Friston, 2005; Rao & Ballard, 1999). Activation in higher-level visual areas represent the generated prediction based on lower-level input, while lower-level responses represent the mismatch between sensory and predicted input (i.e., prediction error). Predictive signals from higher-level areas are sent back to lower-level areas to reduce prediction error by suppressing sensory signals that can be expected based on the higher-level hypothesis. Presumably, early visual areas receive inhibitory feedback from the motion areas hMT/V5+ (Ahmed et al., 2008; Vetter et al., 2015; Wibral et al., 2009). Several physiological studies have indeed demonstrated that sensory signals which can be predicted from their surrounding motion context evoke smaller responses in V1 (Alink, Schwiedrzik, Kohler, Singer, & Muckli, 2010; Harrison, Stephan, Rees, & Friston, 2007; Schellekens, van Wezel, Petridou, Ramsey, & Raemaekers, 2014). AM stimuli may provide such a predictable motion context. A stimulus presented on the AM path can be considered predictable, as it is consistent with the spatio-temporal prediction of a single stimulus in motion. AM masking may therefore be the result of suppression of lower-level responses to the predictable target stimulus. According to this view, AM causes suppression instead of activation in areas such as V1.

In this study, we use computational modelling to uncover which of these two mechanisms actually underlies AM masking, thereby revealing the effects of AM on lower levels of the visual system. Similar to Hidaka et al. (2011), we used grating inducers to create an AM percept and find masking of a target grating along the AM path. If AM masking is indeed the result of activation in lower-level visual areas such as V1, this masking can be considered to be a special case of pattern masking. In a typical pattern masking study, a target grating has to be detected against a stationary background grating. When the contrast of the background grating is sufficiently high, the target grating is masked. Pattern masking is typically attributed to activation at the level of V1. The background grating

evokes a response in V1 cells, and this response can interfere with the response to the target grating. Contrast normalization models provide an excellent account of pattern masking, explaining how activation in low-level stages of visual processing leads to masking (Carandini, Heeger, & Movshon, 1997; Goris, Wichmann, & Henning, 2009; Watson & Solomon, 1997). If AM indeed induces V1 activation, in line with the early filling-in hypothesis, and if that activation is the sole cause of AM masking, a normalization model should be able to account for this masking.

We developed a V1-like population code model by extending a standard contrast normalization model to include effects of AM. We find that this model cannot account for our results when incorporating only AM-related activation. The amount of activation predicted by the model is too small to be perceived by our observers and does not cause significant masking. Instead, a model incorporating strong suppression of responses to stimuli in the AM path captures the observed masking effects, arguing in favour of predictive coding theory.

4.2 Results¹

4.2.1 AM masking

We measured the effect of perceiving AM on the detection of a target grating in a spatial two-alternative forced-choice (2AFC) task. In the AM condition, two alternately presented grating stimuli induced a strong percept of AM along a vertical path at both sides of the screen. The target grating was presented either at the left or right side, in the middle of the path. In a Flicker control condition, the two inducers appeared simultaneously, which disrupted the motion percept entirely. Grating contrast and the difference between target and inducer orientation were varied systematically (see Figure 4.1A for an illustration of a trial sequence in the AM condition).

When the target orientation matched the orientation of the inducers, strong masking was

¹In accordance with journal guidelines, the Methods section of this chapter will be provided at the end of the chapter.

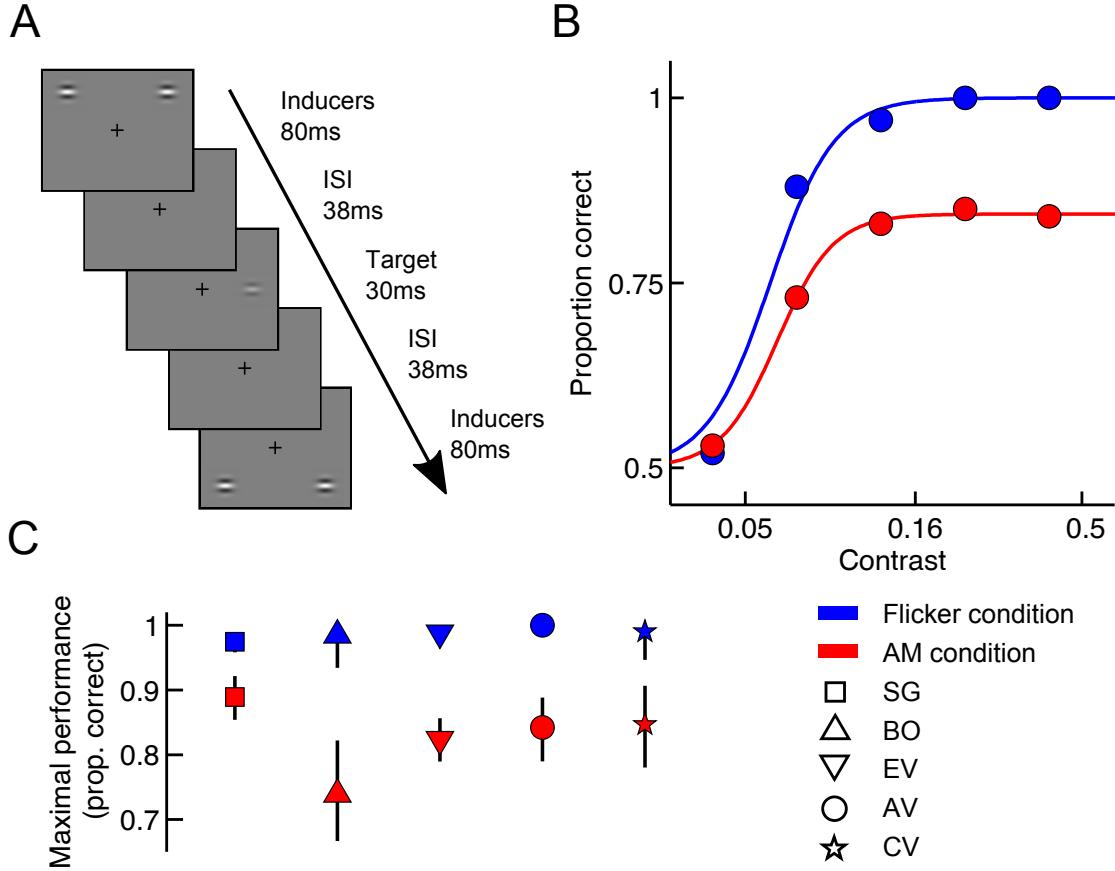


Figure 4.1. AM induces decreased contrast detection performance. (A) Illustration of the stimulus sequence in the AM condition of the 2AFC task. The contrast level of the target grating has been increased for illustration purposes. (B) Maximum contrast detection performance of a typical observer (AV) is considerably lower in the AM condition (red) than in the Flicker condition (blue). Full lines depict the best-fitting logistic psychometric function. (C) Maximal detection performance ($1 - \lambda$) is lower in the AM condition than in the Flicker condition for all observers. Symbols denote different observers. Error bars represent the 95% confidence interval.

observed during AM. Observers failed to reach a high detection performance in the AM condition, even when grating contrast was high. The maximum performance is captured by the psychometric function via $1 - \lambda$ (see Methods) and was estimated at 83% on average across observers (Figure 4.1C). The difference in maximum performance between AM and Flicker condition was significant (average λ difference = 16%, parametric bootstrap, $p < 0.05$ for all observers after Bonferroni correction). The position of the psychometric function along the contrast axis did not significantly differ between the AM and Flicker condition. Figure 4.1B shows the data of a typical observer when target orientation was identical to the orientation of the inducers. As all observers displayed similar patterns,

we pooled the data across observers (see Appendix A in section 4.5). The pooled data set will be used in the remainder of this study.

Figure 4.2 displays the psychometric functions fitted to the pooled data for each orientation level. AM masking appears to be tuned for orientation: the observed masking in the AM condition decreased when the orientation difference between target and inducers increased. This is also evident from Figure 4.3, which shows the maximum performance level ($1 - \lambda$) as a function of the orientation difference for the AM and Flicker condition. Maximum performance ($1 - \lambda$) in the AM condition increased significantly when the difference between target and inducer orientation was increased from 0° to 45° (linear regression slope = 0.0028, parametric bootstrap, $p < 0.001$). In the Flicker condition however, performance was constant in the 0° - 45° range (linear regression slope = -0.0001, parametric bootstrap, $p = 0.8645$). Increasing the orientation difference from 45° to 90° did not affect maximal performance in either the AM or Flicker condition (pairwise comparison, parametric bootstrap, $p > 0.05$ for both conditions). Maximum performance was significantly higher in the Flicker condition than in the AM condition for all orientation levels (pairwise comparison, parametric bootstrap, $p < 0.05$ after Bonferroni correction). The steepness of the psychometric function controlled by s did not differ between the AM and Flicker condition for any of the orientation levels (average difference = 0.0025, parametric bootstrap, $p > 0.05$ for all observers).

4.2.2 Population code model of AM-induced effects

We developed a V1-like population code model (see Methods), based on the contrast normalization model (Carandini et al., 1997; Goris, Putzeys, Wagemans, & Wichmann, 2013; Putzeys, Bethge, Wichmann, Wagemans, & Goris, 2012; Watson & Solomon, 1997), to incorporate AM-induced effects on low-level visual processing. In our model, the effects of AM on the encoding of a target grating can potentially occur in three different ways (for a schematic overview, see Figure 4.4). First, AM can cause excitation by “filling-in” activation along the AM path. More specifically, AM can induce a response in neurons sensitive to the inducer orientation. Second, AM can inhibit responses in the neural

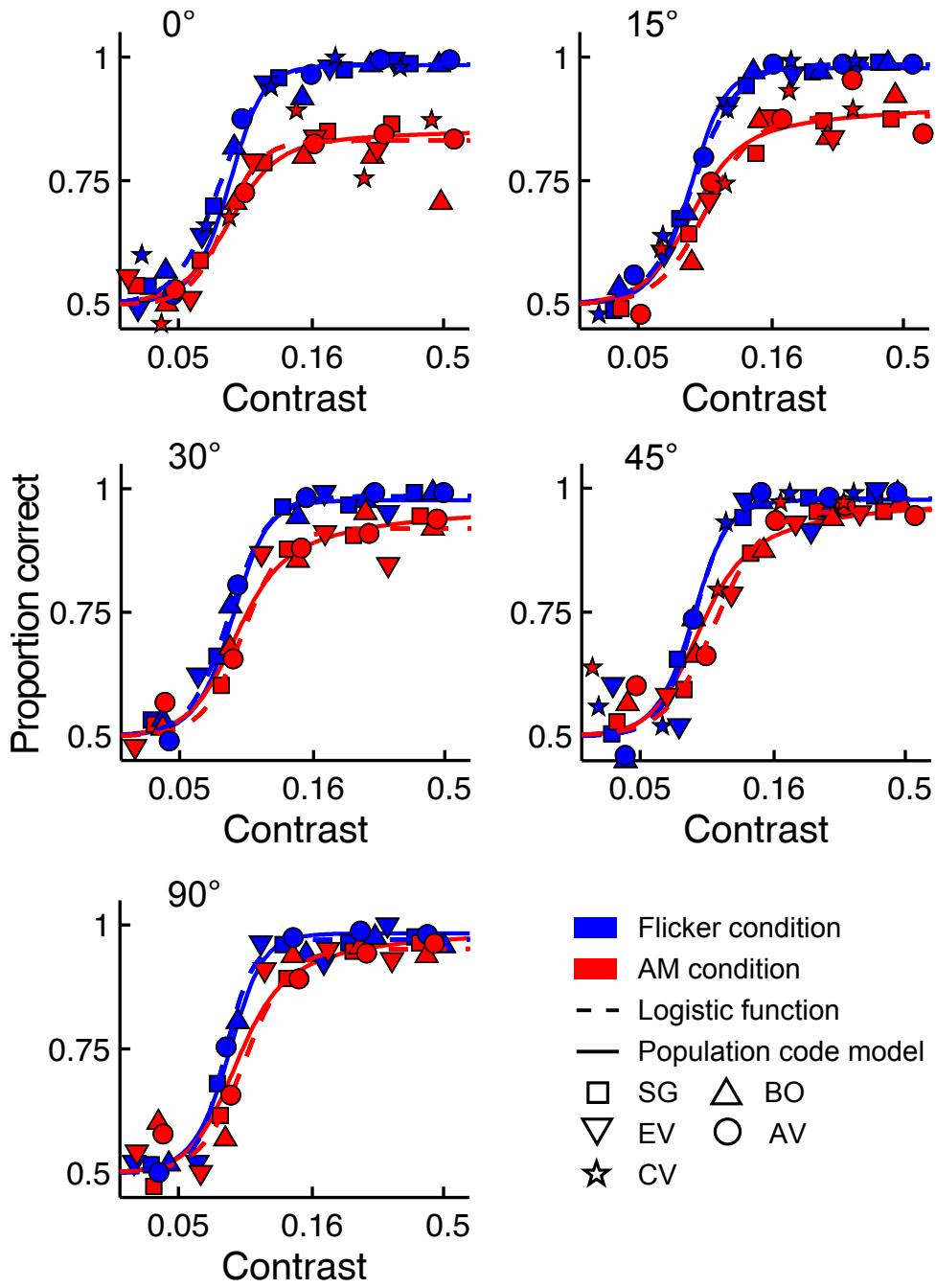


Figure 4.2. Orientation tuning of AM masking and model fits. The pooled data of the five observers are shown for each orientation level. Red and blue symbols represent the AM and Flicker conditions, respectively. Dashed lines depict the best-fitting logistic psychometric function, while full lines represent the best-fitting contrast normalization model. Symbols denote different observers.

population by exciting the gain-control pool and hence causing normalization. It should be noted that the excitatory and inhibitory AM effects mimic the excitation and inhibition that would be observed in the case of real motion. In other words, the neural population

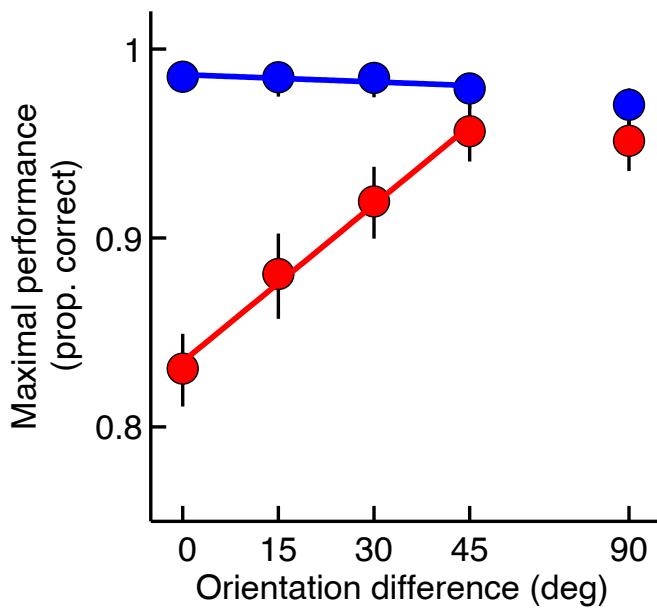


Figure 4.3. Dependence of maximal performance on orientation of the target grating. Full lines depict regression lines reflecting maximal performance for orientation differences between target and inducers in the $0^\circ - 45^\circ$ range. Error bars denote the 95% confidence interval. When orientation difference increased from 0° to 45° , maximal performance ($1 - \lambda$) increased significantly in the AM condition (red), while performance remained constant in the Flicker condition (blue). Maximal performance was not affected when increasing the orientation difference from 45° to 90° in either the AM or Flicker condition.

responds “as if” the inducer was actually moving along the AM path, in accordance with a “filling-in” account of AM. A third way AM can affect the population response is by reducing the response gain of neurons sensitive to the inducer orientation. In this case, the maximum response is reduced as the entire contrast response function is rescaled to lower response rates. We consider such a suppressive effect because of our finding that AM reduces maximum performance and that a rescaled psychometric function is required to capture this observation. Note that, in this study, AM-induced effects refer to effects caused by the inducers during the AM sequence, presumably via feedback from motion areas, and not necessarily caused by the conscious percept of AM.

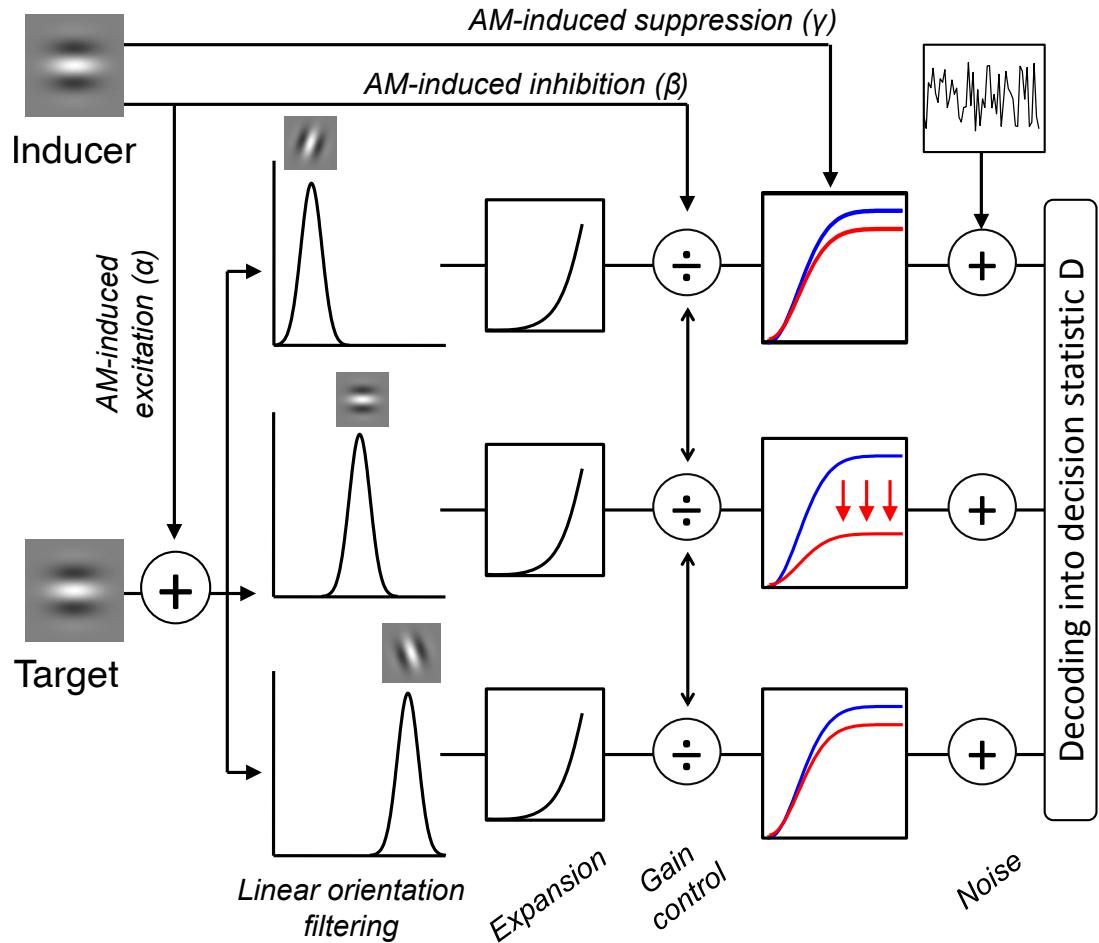


Figure 4.4. Schematic overview of possible AM-induced effects in the contrast normalization model. Following the standard contrast normalization model, the target grating in our task is encoded by a population of V1-like neurons tuned to orientation, which are subject to response acceleration and divisive inhibition. This standard model is extended by including the effects of AM, which can modify the encoding of gratings in our model in three ways. (1) AM can excite linear receptive fields sensitive to the orientation of the inducers controlled by α , evoking responses as if the inducer was physically present at the target location (as during real motion). This would correspond to a “filling-in” process along the AM path. (2) AM can induce divisive normalization via β by exciting neurons in the gain-control pool tuned to the orientation of the inducers. A similar divisive normalization signal would be observed when the inducer would be positioned at the target location. As such, AM-induced inhibition is also in accordance with a “filling-in” account of AM. (3) AM can scale down the contrast response functions due to the suppressive effect exerted by neurons tuned to the inducers’ orientation via γ .

4.2.3 Evaluation of model fit

Our model accurately captures the observed AM masking and its dependence on the orientation difference between the target and inducers (see Figure 4.2). We compared the AIC of this model with the AIC of the best-fitting logistic psychometric function

model which depends on fewer theoretical constraints. The AIC of our model was not significantly higher than this model (AIC difference = 12.74, parametric bootstrap, $p = 0.27$), meaning that our population code model provided a good fit in comparison to a highly flexible psychometric function model.

4.2.4 Response suppression during AM

Maximum-likelihood fitting provided the optimal parameter estimates of our model. Most estimates are well within the range of values reported in physiological V1 single-cell recording studies in monkeys or cats. The semi-saturation contrast was estimated at 9.65%. A similar value has been reported in a physiologically plausible population code model of human contrast processing (Chirimuuta & Tolhurst, 2005). The spontaneous background activity equalled 4.55% of the maximal response, matching reports of Geisler and Albrecht (1997). The concentration parameter k_{exc} , controlling the bandwidth of excitatory (Von Mises) orientation tuning functions, was estimated at 1.35. This corresponds to a half-width orientation bandwidth at half-height of 20.96°, in agreement with observed bandwidths of V1 orientation tuning functions (De Valois, Yund, & Hepler, 1982). The response exponent p , determining the degree of non-linear response expansion, was estimated at 5.52. This is a relatively high value in comparison with physiological findings (see Methods and Discussion; Heeger, 1992). Our efficiency parameter ϵ was estimated at 67%. As noted previously, the exact value of the efficiency parameter may reflect a wide range of possible factors which merely affect absolute performance and do not mediate AM-specific masking.

Possible effects underlying AM in our model are excitation, inhibition through divisive normalization and response gain suppression, controlled by α , β , and γ , respectively. AM induces significant excitation: α was estimated at 1.17% (95%CI = [0.62%, 1.68%]). AM also causes significant inhibition, as estimated by β which equalled 1.51% (95%CI = [1.20%, 2.06%]). Importantly, AM-induced excitation and inhibition have only a limited effect on the contrast response functions and, consequently, on detection performance. To demonstrate the contribution of excitation and inhibition to masking, Figure 4.5

shows the predictions of the model for the case in which α and β were set to zero after model fitting. It can be seen that the model still predicts a considerable amount of masking when AM-induced excitation and inhibition are removed. The reason is that masking is mainly caused by response gain suppression in our model. γ equalled 64.4% (95%CI = [56.4%, 76.5%]), indicating that the contrast response function in the presence of AM is scaled down by a factor of $(100 - 64.4)\% = 35.6\%$. Figure 4.5 shows the predictions of the best fitting model in which γ was set to zero after fitting. Not only is masking significantly reduced at high contrast levels, at low contrast levels the model predicts facilitation. In other words, detection performance is predicted to be better in the AM condition compared to the Flicker condition when target grating contrast is low. This facilitation effect was not present in the data. In agreement with the predictions of the logistic psychometric function model, the population code model predicts that performance in the AM condition is always lower than in the Flicker condition, also at lower contrast levels. To further test whether a model including only excitation and inhibition could account for our results, we refitted the model without suppression (γ was constrained to zero before fitting). The AIC of this model was significantly worse than the AIC of our full model including suppression (AIC difference = 212.62, parametric bootstrap, $p < 0.001$), indicating that the latter provided a better fit to the data.

4.2.5 AM-induced effects are tuned to orientation

The effects induced by AM in our model are selective to orientation. More specifically, the size of the AM effects on a neuron decreases as the difference between the neuron's preferred orientation and the inducer orientation increases. For AM-induced excitation as well as response gain suppression, the tuning functions have an orientation bandwidth equal to the bandwidth of the neurons' excitatory receptive fields ($k_{exc,AM} = k_{exc} = 1.35$). AM thus only affects neurons sensitive to the inducer orientation, with an orientation precision that matches the orientation selectivity of typical V1 cells (De Valois & De Valois, 1988). The bandwidth concentration parameter of the tuning function for the AM-induced inhibition was estimated at 0.001. This small value implies that inhibition is not tuned for the inducer orientation and is consistent with a broadly tuned gain control

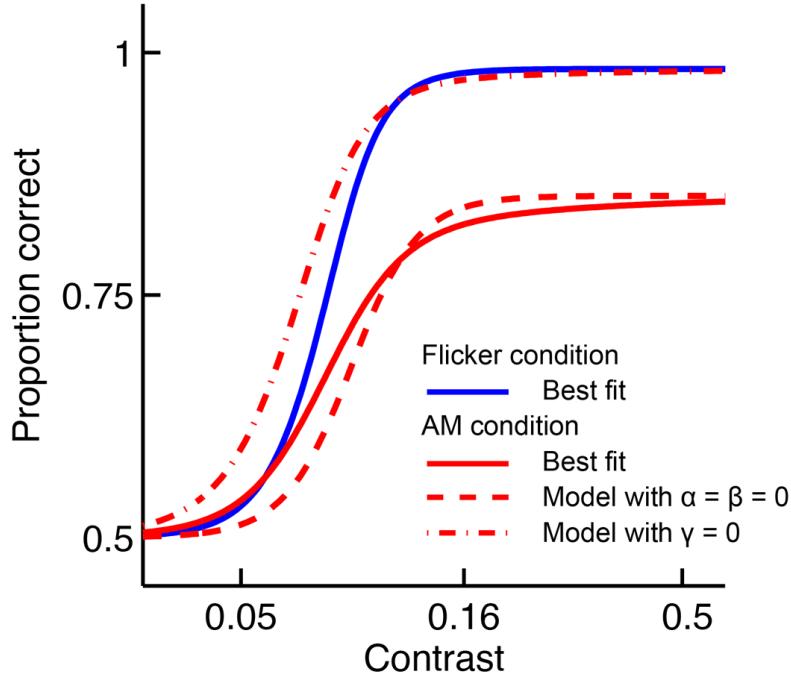


Figure 4.5. Comparison of contrast normalization models in capturing AM masking. The pooled data of five observers are shown for the condition in which the orientation of target and inducers is identical. The predictions of the best-fitting contrast normalization model including AM-induced effects is indicated by the full red line (AM condition) and full blue line (Flicker condition). The dashed line shows the model prediction for the AM condition when excitation and divisive inhibition effects are removed after fitting ($\alpha = \beta = 0$). The dash-dotted line represents the prediction when the suppression effect is set to zero after fitting ($\gamma = 0$). In contrast to the model without excitation and divisive inhibition, the model without suppression fails to account for the observed masking at high contrast levels. In addition, this model predicts facilitation of contrast detection at low contrast levels, which is not supported by the data. All three models predict the same performance for the Flicker condition, as all parameters controlling AM effects (α , β and γ) are set to zero for this condition.

pool. To examine whether a model which allows no orientation tuning of the AM-evoked effects could also account for our data, we compared the AIC of our model with the AIC of a model in which $k_{exc,AM}$ and $k_{inh,AM}$ are fixed at 0.001 before fitting. Our full model was evaluated significantly better than the model containing no tuning of AM effects (AIC difference = 136.55, parametric bootstrap, $p < 0.001$).

4.2.6 Early excitation is not the cause of perceptual filling-in

The small level of AM-induced excitation can be seen as filling-in of activation along the AM path, as mentioned earlier, but to what extent can it be considered as *perceptual*

filling-in? To answer this question, we use our population code model to predict the detectability of a grating evoking the same level of activation as induced by AM. This is a grating with contrast equal to $\alpha = 1.17\%$, following from the fact that the estimated values of k_{exc} and $k_{exc,AM}$ are equal (see Equation 4.10 in Methods section). Our best-fitting model predicts that even in the Flicker condition, in which the AM-induced suppression is absent, such a grating cannot be detected (detection performance would equal 50%).

4.3 Discussion

In the present study, we examined the underlying neural mechanism of AM in low-level visual processing. To this end, we investigated how AM influences the detection of grating stimuli presented along the AM path. We found that AM impairs the detection of a target presented on the AM path when the target's orientation matches the inducers' orientation. Other studies reporting AM masking typically measured detection performance at a single stimulus intensity level. We evaluated detection performance across a wide range of grating contrast levels in order to obtain full psychometric functions. This allowed us to discover that perceiving AM imposes an upper bound on performance, which cannot be exceeded even when grating contrast is raised to a substantial level.

We applied a V1-like population code model, based on contrast normalization, and found that this model provides an excellent account of the data, predicting the AM-induced upper bound on performance. Importantly, the model reveals that AM masking is not caused by excitation but by suppression of responses to stimuli along the AM path, through a reduction in response gain. In fact, when this suppressive effect is removed from our model, masking disappears and the model even predicts facilitation. This is related to the well-known pedestal effect, which refers to the observation that the presence of a background grating can facilitate the detection of a superimposed target grating at low background contrasts (Nachmias & Sansbury, 1974). The pedestal effect has been attributed to the fact that the background grating evokes a small level of activation in V1 (Goris et al., 2009). Due to this activation, target detection operates on a steeper part

of the V1 contrast response functions, leading to larger differential responses. According to our model, the AM-induced activation also acts as a pedestal, improving detection performance. This facilitation effect does not show up in the data due to the strong suppressive effect of AM. Moreover, we showed that a grating on the AM path evoking the same level of excitation as predicted by our model would not be detectable. The AM-evoked excitation can thus hardly be considered as an instance of perceptual filling-in. This provides strong support for the hypothesis that perceptual filling-in of the AM path occurs at a later processing stage, presumably in motion area MT+ (Liu et al., 2004).

A large response exponent had to be assumed in the model to account for the large steepness of the psychometric functions. This response exponent estimate does not only reflect the degree of V1 response acceleration in our model (see Methods). The parameter also captures other factors contributing to psychometric function steepness which were not incorporated in our model to keep computations tractable. Spatial uncertainty, for instance, has been found to increase the steepness of the psychometric function (Pelli, 1985). Uncertainty about the target location has been found to increase with eccentricity (Michel & Geisler, 2011; Najemnik & Geisler, 2005). In our experiment, the target was positioned relatively far in the periphery (i.e., 10°), making it possible that uncertainty contributed to the steepness of the psychometric functions. Importantly, it is unlikely that spatial uncertainty played any role in AM, as the slope of the psychometric functions did not differ between the AM and Flicker condition. A reduction in maximum performance, which is the key aspect of AM masking, has not been linked to spatial uncertainty (Michel & Geisler, 2011).

The response gain reduction predicted by our model was strongest when the target matches the orientation of the motion-inducing stimuli. This orientation tuning of AM-induced suppression is implemented in the model by rescaling the effects of AM on a given neuron with the selectivity of that neuron to the inducers. AM thus only affects neurons that are selective for the inducers' orientation. The bandwidth of the tuning functions capturing this selectivity match physiological estimates of V1 orientation bandwidth (De Valois & De Valois, 1988). This strongly supports our claim that AM-induced

suppression occurs in lower-level visual areas such as V1. In addition, the observed orientation tuning makes it unlikely that AM masking can be explained merely by increased overt or covert spatial attention to the inducers in the AM condition compared to the inducers in the Flicker condition. Such spatial attention to the inducers should be independent of stimulus features such as grating orientation presented at the unattended target location.

Our results of suppressed activation fit a predictive coding scheme (Friston, 2005; Rao & Ballard, 1999). Responses in low-level areas signaling prediction error are suppressed when they are consistent with the higher-level prediction of motion by a single stimulus along the AM path, in accordance with other studies reporting reduced V1 activation for local features that fit their surrounding context (Alink et al., 2010; Harrison et al., 2007; Fang, Kersten, & Murray, 2008; Murray, Kersten, Olshausen, Schrater, & Woods, 2002; Schellekens et al., 2014). For instance, Alink et al. (2010) found that the predictability of stimuli in their surrounding AM context leads to reduced activation in V1. However, since the authors did not include a control Flicker condition, responses to a predictable stimulus in the context of AM could not be compared to those evoked by an unpredictable stimulus. Our model predicts that responses to a physically present stimulus in the AM condition would be lower than those to a stimulus in the Flicker condition and that this suppression of responses to predictable sensory input is the main cause of AM masking. It should be noted that this is different from the paradigm used by Muckli et al. (2005), in which no stimulus was present along the AM path and suppression of responses to such a stimulus could consequently not be measured. Suppression through response gain reduction is a multiplicative effect that can be implemented in a predictive coding framework through backward cortico-cortical connections (Rao & Ballard, 1999). Feedback from higher-level neurons has been shown to modulate bottom-up responses of lower-level neurons in a multiplicative manner (Herrmann, Montaser-Kouhsari, Carrasco, & Heeger, 2010; Webb, Tinsley, Barraclough, Parker, & Derrington, 2003; Williford & Maunsell, 2006). In addition, it has been found that GABA_A inhibition controls response gain in V1, without affecting contrast gain (Katzner, Busse, & Carandini, 2011). It is possible that GABA_A levels are selectively increased during AM. Indeed, GABA-mediated

cortical inhibition has been linked to predictive coding in a recent study (Matthews et al., 2014).

In our model, V1 responses are inhibited during AM in two distinct ways, namely reduction in response gain and contrast gain (i.e., contrast sensitivity) (Chaumon & Busch, 2014; Herrmann et al., 2010; Webb et al., 2003). Contrast response functions were found to be rescaled to lower response rates during AM, indicating a decrease in response gain. At the same time, a gain control mechanism is activated during AM, which shifts the contrast response function to higher contrasts, thereby lowering contrast sensitivity. Importantly, the response gain effect is much larger than the change in contrast sensitivity. In addition, the change in response gain is much more narrowly tuned for the inducers' orientation. This seems to suggest that the two forms of inhibition reflect different neural mechanisms. We believe that the response gain reduction is a direct consequence of AM because it is narrowly tuned. The change in contrast sensitivity seems to be a more indirect effect in that it may result from the activation of the gain-control mechanism by the small AM-induced excitation. Indeed, it has often been found that the gain control mechanism is broadly tuned for orientation (Foley, 1994). Presumably, inhibition by neurons that are not sensitive to the inducers' orientation results in weak AM masking to be still present in our data when the target and inducers are orthogonal. We incorporated response gain reduction in our model as a linear rescaling of the contrast response function similar to previous studies (Herrmann et al., 2010; Webb et al., 2003; Williford & Maunsell, 2006), but other implementations may be possible. For instance, Rosenberg et al. implement a decrease in response gain via divisive normalization (Rosenberg, Patterson, & Angelaki, 2015). More specifically, they assume an amplification of the normalization signal, which causes the contrast response functions to scale to lower response rates. This non-linear rescaling approximates the linear rescaling implemented in our model to a large degree. It is therefore possible that a model incorporating changes in divisive normalization during AM may account for the observed suppression. However, such a model would also have to predict the contrast gain reduction observed during AM in the present study. A contrast gain decrease corresponds to an additive increase of the normalization signal, whereas response gain reduction is realized as through a multiplica-

tive increase. Although it is possible that changes in the normalization mechanism are compound, consisting of both additive and multiplicative effects, previous studies have mainly reported additive changes leading to a change in contrast gain (Carandini et al., 1997). Our model therefore only incorporates additive changes in normalization during AM.

In the present study, we made three important contributions to the understanding of AM. First, we discovered a central but hitherto unnoticed aspect of AM masking, namely the upper bound on detection performance. Second, we identified orientation-tuned suppression of V1 responses as the major cause of AM masking. Third, we concluded that perceptual filling-in of the motion path does not occur at an early stage of visual processing. Our results are consistent with predictive coding models of cortical processing, proposing that higher-level predictions of motion are generated which “explain away” lower-level responses to expected sensory input. Further research is needed to determine whether predictive coding is able to provide a coherent account of how AM is implemented in the brain.

4.4 Materials and Methods

4.4.1 Subjects

Five observers (AV, BO, CV, EV, and SG) participated in the experiment. All were naive to the purpose of the study and had normal or corrected-to-normal vision (age range 20-23). The study was approved by the Social and Societal Ethics Committee of the University of Leuven. Written consent was obtained for all participants before the start of the experiment. Observers were paid 8 euros an hour for participating. A block of 50 practice trials were conducted to familiarize subjects with the stimuli and task. All subjects reported having no difficulty in perceiving the AM sequences.

4.4.2 Apparatus

Stimuli were generated and presented using a ViSaGe stimulus generator (Cambridge Research Systems, Cambridge, England) controlled by MATLAB (MathWorks, Natick, US). A linearized ViewSonic G90fB monitor (ViewSonic, California, USA) was used to display the stimuli. The monitor had a spatial resolution of 1024 x 768 pixels and operated at a refresh rate of 118 Hz, with 8-bit luminance precision for all contrast levels used in the study. Participants were seated in a darkened room with their heads supported by a chin rest, at a viewing distance of 60 cm (corresponding to a pixel size of 0.0315° of visual angle). The mean background luminance of the screen was equal to 72.5 cd/m². Participants' responses were registered by means of a Cedrus response box (RB-530, Cambridge Research Systems).

4.4.3 Stimuli

All stimuli used in the experiment were Gabor patches, created by multiplying a cosine grating with a 2D Gaussian envelope (SD = 0.75°). The spatial frequency of all gratings was 1.5 cycles per degree. Stimuli were displayed on a gray background (Michelson contrast of 50%). The AM-inducing stimuli had a Michelson contrast of 100%, while the contrast levels of the target stimulus ranged from 4% to 40% Michelson contrast. Target orientation equalled 0° (horizontal), 15°, 30°, 45° or 90° (vertical). The orientation of the inducers was 0°. The target grating was presented at 10° eccentricity either left or right from a fixation cross (0.76° x 0.76°). The AM-inducing stimuli were vertically separated by 8° and the target stimulus was positioned exactly in between, at a distance of 4° from each inducer stimulus.

4.4.4 Detection task

Observers were instructed to detect a target stimulus appearing either at the left or right of a fixation cross in a spatial two-alternative forced-choice (2AFC) task. At the beginning of the trial, a fixation cross was shown for 500 ms. The inducers were then presented for

a duration of 80 ms alternately at the top and bottom position with an inter-stimulus interval of 106 ms. This AM sequence was repeated four times to induce a strong percept of stimuli moving back and forth. The target was flashed briefly for 30.8 ms during the fourth AM sequence, 38 ms after the presentation of the inducer at the top position and at an intermediate position in the interstimulus interval between the two inducers. At the end of the trial, observers were asked to indicate at which side of the screen the target appeared. Auditory feedback was provided after each trial. In the Flicker condition, the procedure was identical, except that the inducers simultaneously appeared at the top and bottom. This disrupted the percept of motion completely, while controlling for masking effects resulting from the presence of the inducers (Kanai & Kamitani, 2003). Figure 4.1A shows an example of the stimulus sequence in the AM condition in which the orientation of the target and inducers are the same.

The experiment consisted of blocks of 50 trials, in which contrast (5), orientation (5) and condition (2) levels were randomised. Each subject completed at least 50 trials for each combination of these levels. Due to time constraints, subject CV only completed three of the orientation levels, namely 0°, 15°, and 45°. Psychometric functions were fitted to the individual and pooled data, relating target grating contrast to proportion correct responses. The form of our psychometric function is given by:

$$\psi(c; c_m, s, \lambda) = 0.5 + (0.5 - \lambda)F(c; c_m, s) \quad (4.1)$$

where c denotes the contrast levels of the target grating and F is a sigmoidal logistic function of c ranging from 0 to 1 (Wichmann & Hill, 2001a):

$$F = \frac{1}{1 + e^{(c_m - c)/s}} \quad (4.2)$$

in which c_m equals the midpoint contrast and s determines the steepness of the function. λ controls the upper bound of the psychometric function, as ψ ranges from 50% to a maximum of $1 - \lambda$. Note that λ is often considered a lapse rate parameter, reflecting the amount of stimulus-independent errors made by the observer (Wichmann & Hill, 2001a). However, λ is estimated here for both the AM and Flicker condition and for each orienta-

tion of the target. Hence, λ should not be interpreted as a lapse rate, as it will be evident from our results that λ is highly dependent on these stimulus conditions. Psychometric functions were fitted using a maximum-likelihood fitting procedure (Wichmann & Hill, 2001a). A parametric Monte-Carlo bootstrap procedure involving 10,000 samples provided the distributions of the deviance statistic used to assess goodness-of-fit, as well as the confidence intervals for the parameter estimates (Wichmann & Hill, 2001b). It should be noted that the goodness-of-fit of all psychometric functions fitted in the present study was acceptable (parametric bootstrap, $p > 0.05$ after Bonferroni correction).

4.4.5 Population code model

Encoding stage. Gratings in our task are encoded by a population of V1-like neurons. Each neuron in the population is characterized by a linear excitatory receptive field tuned to orientation (Hubel & Wiesel, 1959). The response of this receptive field to a grating of contrast c and orientation θ is provided by:

$$L_i(c, \theta) = c * f_i(\theta) \quad (4.3)$$

where f is a von Mises orientation tuning function (Swindale, 1998) rescaled to obtain a maximum of one at the preferred orientation $\hat{\theta}_i$, irrespective of the bandwidth of the function:

$$f_i(\theta) = \frac{e^{k_{exc}(2\cos(2(\theta - \hat{\theta}_i)) - 1)}}{e^{k_{exc}}} \quad (4.4)$$

The bandwidth is controlled by the concentration parameter k_{exc} . Note that we implicitly assume that each neuron in the population is spatially tuned to the location of the target grating in the middle of the AM path. Given the large spatial separation of the target and inducers, the neurons in our population are further assumed to operate independently from neurons tuned to the inducer locations. Hence, in our model, neurons tuned to other spatial locations do not play a role in the detection of the target.

The linear response $L_i(c, \theta)$ is raised to an exponent p to introduce an accelerating non-linearity (Heeger, 1992), and divided by a normalization term according to the Naka-

Rushton equation (Albrecht & Hamilton, 1982; Naka & Rushton, 1966):

$$R_i(c, \theta) = \frac{L_i(c, \theta)^p}{c_{50}^p + G_i(c, \theta)^p} \quad (4.5)$$

$G_i(c, \theta)$ is the normalization signal, i.e., the linear response of a divisive inhibitory contrast gain control pool, defined as:

$$G_i(c, \theta) = c * g_i(\theta) \quad (4.6)$$

where g captures the orientation tuning function of the gain-control pool (Bonds, 1989). g_i is identical to f_i , except that a different bandwidth parameter k_{inh} is used. The responses of neurons in the gain-control pool inhibit the response of neuron i to the target grating, thereby causing the response R_i of this neuron to saturate at contrasts above c_{50} , which is also known as the semi-saturation contrast. The average response \bar{r}_i to the target grating (in number of spikes) is obtained by extending Equation 4.5 to incorporate spontaneous discharge r_0 (in Hertz), maximum response rate r_{max} (in Hertz) and stimulus presentation duration t (in seconds):

$$\bar{r}_i(c, \theta) = t \left[r_0 + r_{max} \frac{L_i(c, \theta)^p}{c_{50}^p + G_i(c, \theta)^p} \right] \quad (4.7)$$

Response variance is proportional to the average response rate (Vogels, Spileers, & Orban, 1989):

$$var(r_i) = \zeta \bar{r}_i \quad (4.8)$$

where ζ is a proportionality constant. Individual neural responses r_i are assumed to follow a normal distribution:

$$r_i \sim N[\bar{r}_i, var(r_i)] \quad (4.9)$$

Implementing the effect of response covariance requires elaborate Monte-Carlo simulations. To keep computations tractable, a diagonal covariance matrix is used, thereby assuming that correlations between neural responses are zero. However, V1 neurons are known to be correlated (Cohen & Newsome, 2009; Zohary, Shadlen, & Newsome, 1994). Note that these correlations only scale down the average signal-to-noise ratio of the population response. A lower overall signal-to-noise ratio results in lower average detection

performance across all conditions. To capture such variations in overall detection performance, we included a global efficiency parameter in the model’s decoding stage (cf. infra).

Equation 4.7 defines the contrast response function of the standard contrast normalization model (Carandini et al., 1997; Goris et al., 2013; Putzeys et al., 2012; Watson & Solomon, 1997). This model is used to predict responses when AM is absent, i.e., in the Flicker condition. To account for the effects of AM, however, Equation 4.7 has to be extended. The average response of neuron i to the target grating in the presence of AM is given by:

$$\bar{r}_i(c, \theta_{tgt}, \theta_{ind}) = t \left[r_0 + [1 - \gamma * h_i(\theta_{ind})] r_{max} \frac{[L_i(c, \theta_{tgt}) + \alpha * h_i(\theta_{ind})]^p}{c_{50}^p + [G_i(c, \theta_{tgt}) + \beta * j_i(\theta_{ind})]^p} \right] \quad (4.10)$$

where h_i and j_i are identical to f_i , except that a bandwidth parameter $k_{exc,AM}$ is used for h_i and a bandwidth parameter $k_{inh,AM}$ is used for j_i . θ_{tgt} represents the target grating orientation and θ_{ind} equals the orientation of the two AM inducing gratings used to create the percept of AM. We assume AM can affect the encoding of gratings by changing the contrast response function in three major ways. (1) AM can introduce a level of excitation, controlled by α , (2) AM can cause inhibition, i.e., a shift of the contrast response function towards higher contrasts, controlled by β , and (3) AM can cause suppression, i.e., a reduction of the maximal contrast response, controlled by γ . In the absence of AM (as in the Flicker condition), α , β and γ are zero and Equation 4.10 reduces to Equation 4.7. We now discuss these effects in detail.

Via α , AM induces a neural response provided that the tuning function h_i evaluated at the inducer grating orientation is not zero. AM thus only excites neurons that are sensitive to the inducer orientation. In this way, we obtain responses “as if” the inducer was physically moving along the AM path. These responses thus reflect the “filling-in” of activation. In the special case that $\theta_{tgt} = \theta_{ind}$ and $k_{exc} = k_{exc,AM}$, AM will induce a response that is equal to the excitatory receptive field response that would be evoked by an inducer grating with contrast α presented at the target location (for instance, during a physical motion along the AM path). As the inducer gratings were presented at 100%

contrast during our experiments, complete filling-in occurs when α equals 1. It should be noted that the tuning bandwidth parameter $k_{exc,AM}$ for the AM-induced excitation was allowed to differ from the bandwidth parameter k_{exc} of the linear receptive field when fitting the model. However, the estimates of these parameters did not differ for the best-fitting model (cf. Results section).

As mentioned earlier, the responses of a given neuron i are normalized by the responses of other neurons in a gain-control pool. By evoking responses in these gain-control neurons, AM may cause divisive normalization. β controls the strength of this inhibitory effect. The term $\beta * j_i(\theta_{ind})$ lowers contrast sensitivity by shifting the contrast response function to higher contrasts, but only of neurons tuned to the inducer orientation. It can be seen as an indirect effect, resulting from a more direct excitatory effect that activates the normalization mechanism. The tuning bandwidth parameter of the inhibitory effect $k_{inh,AM}$ was allowed to be different from the bandwidth parameter k_{inh} of the gain-control pool during fitting but again, these parameters were estimated to be equal for the best-fitting model (cf. Results section).

The third AM-induced effect is suppression through a reduction of response gain. γ rescales the contrast response function to lower response rates, thereby reducing the neurons' maximal response. Similar to the other two AM effects, this suppressive effect only occurs in neurons that are tuned to the inducer orientation. The bandwidth of the tuning function is equal to the bandwidth $k_{exc,AM}$ of the excitatory effect.

Decoding stage. In the decoding stage, neural responses are combined into decisions in our spatial 2AFC detection task. We implemented a simple linear decoder that considers the two spatial locations at which the target grating can appear. For each location, the decoder sums the responses of all neurons tuned to that location. The location yielding the largest summed response is indicated as containing the target grating. S_{tgt} equals the summed responses of those neurons tuned to the location at which the target is presented, whereas S_{blank} equals the summed responses of neurons tuned to the other location. For

the AM conditions, the average values of these sums are defined as:

$$\overline{S_{tgt}}(c, \theta_{tgt}, \theta_{ind}) = \sum_{i=1}^N \overline{r_i}(c, \theta_{tgt}, \theta_{ind}) \quad (4.11)$$

$$\overline{S_{blank}}(c = 0, \theta_{tgt}, \theta_{ind}) = \sum_{i=1}^N \overline{r_i}(\theta_{tgt}, \theta_{ind}) \quad (4.12)$$

where $\overline{r_i}$ is provided by Equation 4.10. For the Flicker condition, $\overline{S_{tgt}}$ and $\overline{S_{blank}}$ are obtained in a similar fashion, but using the standard contrast response function of Equation 4.7 instead of the elaborated version of Equation 4.10. In agreement with Vogels et al. (1989), the variance of the summed responses is provided by:

$$var(S_{tgt}) = \zeta \overline{S_{tgt}} \quad (4.13)$$

and

$$var(S_{blank}) = \zeta \overline{S_{blank}} \quad (4.14)$$

Proportion correct detection p in our task is then provided by the cumulative Gaussian function:

$$p = \int_0^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \quad (4.15)$$

with

$$\mu = \overline{S_{tgt}} - \overline{S_{blank}} \quad (4.16)$$

and

$$\sigma = \frac{1}{\epsilon} \sqrt{var(S_{tgt}) + var(S_{blank})} \quad (4.17)$$

where ϵ is an efficiency parameter. This parameter can accommodate the effect of interneuronal correlation as mentioned previously, but can also capture other factors that may affect overall performance such as global attention level and fatigue. The purpose of our study is not to distinguish between these factors, as they only affect absolute performance and do not cause relative differences in performance between AM and Flicker conditions.

It should be noted that our decoder does not use a-priori knowledge of the target grating orientation when summing filter responses. A more optimal decoder may preferentially weight filters that are tuned to the grating while ignoring filters tuned to other, irrelevant orientations. Such a decoder is not plausible in our experiments, however, as multiple target orientations were randomized across trials. Observers did not know the target orientation at the start of each trial. It would be impossible for them to implement a detection strategy tailored to grating orientation without first detecting the grating. In addition, the same decoding strategy is used in the AM and Flicker conditions. In other words, the decoder does not account for the effects of AM on filter responses. The fact that we observe strong masking indeed suggests that the decoder does not manage to discount or compensate for the AM-induced effects on the population response.

Previous studies have shown that observers are to some extent uncertain about the exact spatial location of the target in grating detection tasks, which increases the slope of the psychometric function (Pelli, 1985). This spatial uncertainty effect could be captured by assuming a more complex decoder in our model, for instance, a non-linear decoder that selects the maximum of all neural responses to obtain the decision variable instead of computing a linear sum (Pelli, 1985). Implementing such a decoder would involve elaborate Monte-Carlo simulations as the distribution of the decision statistic cannot be obtained analytically (cf. Equation 4.11 and Equation 4.13). Furthermore, the population of neurons assumed in our model would have to be expanded considerably to include subpopulations that are sensitive to irrelevant locations. As these operations would render model fitting computationally prohibitive, we did not implement non-linear decoding. This does not imply that our model cannot capture the increased slope of the psychometric function in the presence of spatial uncertainty. An increase of the response exponent p allows for such an increased slope. Consequently, the estimated value of p should not be taken to solely reflect V1 response acceleration but may also capture other factors mediating psychometric function steepness. Separating these factors is not a goal of the present study, as we show in the Results section that psychometric function steepness is unrelated to AM.

4.4.6 Model constraints and fitting

Some parameters were poorly constrained by our data and were fixed to physiologically plausible values. These values can be changed without affecting the conclusions of this study. The concentration parameter of the gain control tuning function k_{inh} was set to 0.001, resulting in a broadly tuned gain control pool (Foley, 1994). r_{max} was fixed at 100 Hz (De Valois, Albrecht, & Thorell, 1982). ζ , the proportionality constant controlling response variance, was fixed at 1.9 (Geisler & Albrecht, 1997; Vogels et al., 1989). As physiological studies suggest rather low spontaneous discharge rates at the level of V1, r_0 was constrained to be smaller than 5% of the maximal response r_{max} . The tuning functions controlling the orientation selectivity of the AM effects were not allowed to be narrower than the orientation tuning functions of the excitatory receptive fields. The reason for this constraint is that effects induced by AM are presumably the result of feedback from higher visual areas specialized in motion, such as hMT/V5+ (Ahmed et al., 2008; Vetter et al., 2015; Wibral et al., 2009). These areas are typically characterized by a lower orientation selectivity compared to V1 (De Valois & De Valois, 1988). It is therefore unlikely that the AM effects are more selective to orientation than V1 cells. The response exponent p was not allowed to be smaller than 2 (Heeger, 1992).

A total of 10 parameters were estimated using a maximum-likelihood fitting procedure (Wichmann & Hill, 2001a). Multiple fits were performed using randomized starting values for each parameter. Akaike's Information Criterion (AIC) was calculated to assess the goodness-of-fit of the models while taking into account the complexity of the model quantified as the number of fitted parameters. Parametric Monte-Carlo bootstrapping involving 1000 samples provided the confidence intervals of the estimated parameters and the distributions of the AIC statistics which were used in evaluating the quality of the model fit (Wichmann & Hill, 2001b).

4.5 Appendix A: Individual data and pooling

Each observer exhibited decreased detection performance in the AM condition compared to the Flicker condition. In addition, the AM masking was orientation tuned for each observer. Supplementary Figures 4.6-S4.10 show the individual data sets for each observer. The best-fitting psychometric functions are denoted by dashed lines, while the best-fitting population code models are depicted by full lines.

All observers showed significantly reduced maximal performance ($1 - \lambda$) compared to the Flicker condition. For each observer except SG, the fit of the full population code model was significantly better than a reduced model including only excitation and inhibition (AIC full model - AIC restricted model < 0 , parametric bootstrap, $p < 0.05$ after Bonferroni correction). For observer SG, the model including suppression still outperforms the reduced model (AIC difference = -20.07). The AM-induced suppression was also significant for this observer ($\gamma = 53.36\%$, parametric bootstrap, $95\%CI = [41.22\%, 69.23\%]$). All observers displayed significant orientation tuning of AM masking. A model assuming no tuning of AM effects (i.e., $k_{exc,AM}$ and $k_{inh,AM}$ are fixed at 0.001) provided a significantly worse fit than our full model for all observers (AIC full model - AIC restricted model < 0 , parametric bootstrap, $p < 0.05$ after Bonferroni correction). As patterns were highly similar across observers, the data were normalized and pooled using the following procedure (Putzeys et al., 2012). For each observer and condition (AM or Flicker condition), we estimated the position c_m of the psychometric function along the contrast axis as well as the level of psychometric function compression λ . Within each condition, we divided all contrasts by c_m and multiplied them by the value of c_m averaged across all observers and conditions. Performance levels p were normalized using the following formula:

$$p_{norm} = 0.5 + (p - 0.5) \frac{1 - \lambda_{avg}}{1 - \lambda} \quad (4.18)$$

where p_{norm} is the normalized performance and λ the compression level for a specific observer and condition. λ_{avg} is the value of λ averaged across observers and conditions. After normalization, data were pooled to obtain a single data set representing the average performance of all observers.

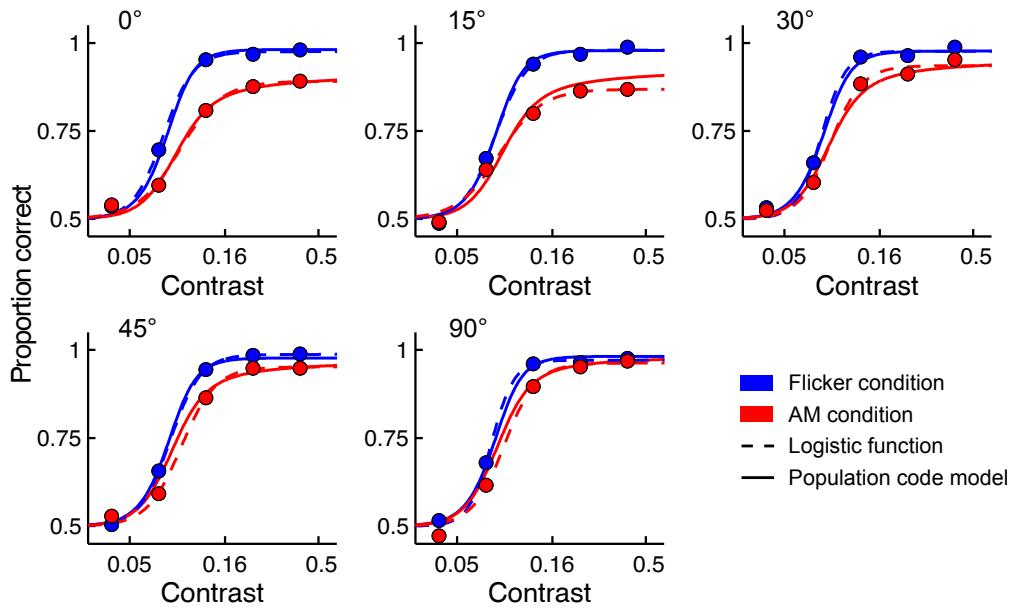


Figure 4.6. Model fits to data of observer SG. Plots are shown for each orientation difference between the target and inducer gratings. Red and blue symbols denote the AM and Flicker conditions respectively. Dashed lines depict the best-fitting logistic psychometric functions, while full lines represent the best-fitting population code model.

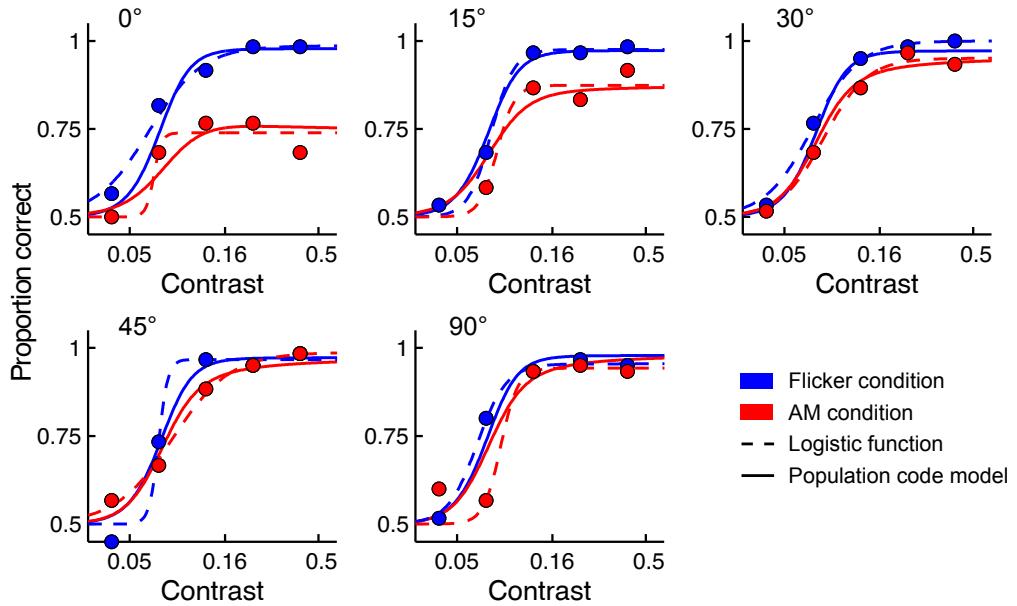


Figure 4.7. Model fits to data of observer BO. Plots are shown for each orientation difference between the target and inducer gratings. Red and blue symbols denote the AM and Flicker conditions respectively. Dashed lines depict the best-fitting logistic psychometric functions, while full lines represent the best-fitting population code model.

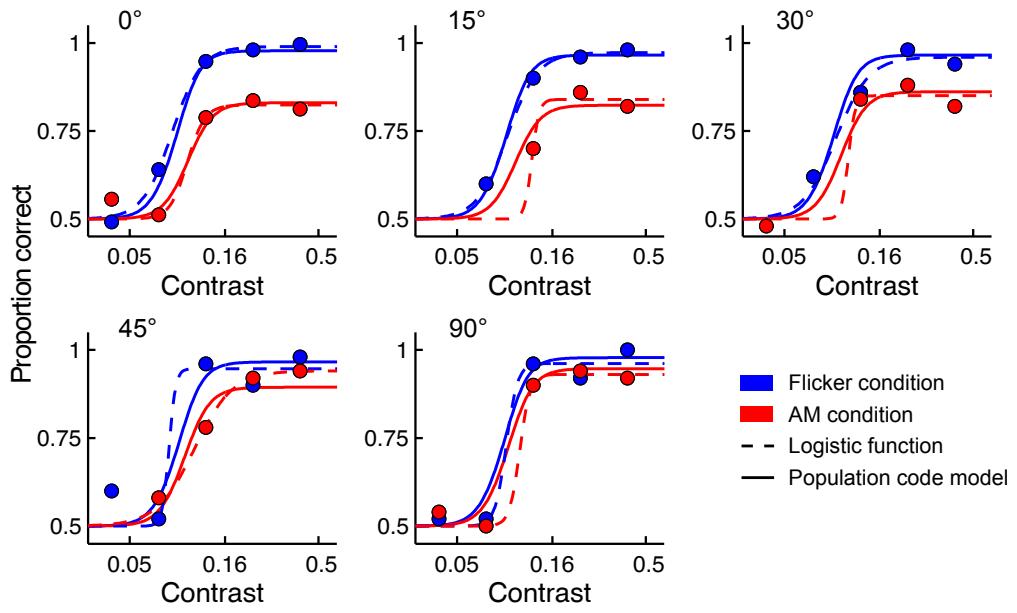


Figure 4.8. Model fits to data of observer EV. Plots are shown for each orientation difference between the target and inducer gratings. Red and blue symbols denote the AM and Flicker conditions respectively. Dashed lines depict the best-fitting logistic psychometric functions, while full lines represent the best-fitting population code model.

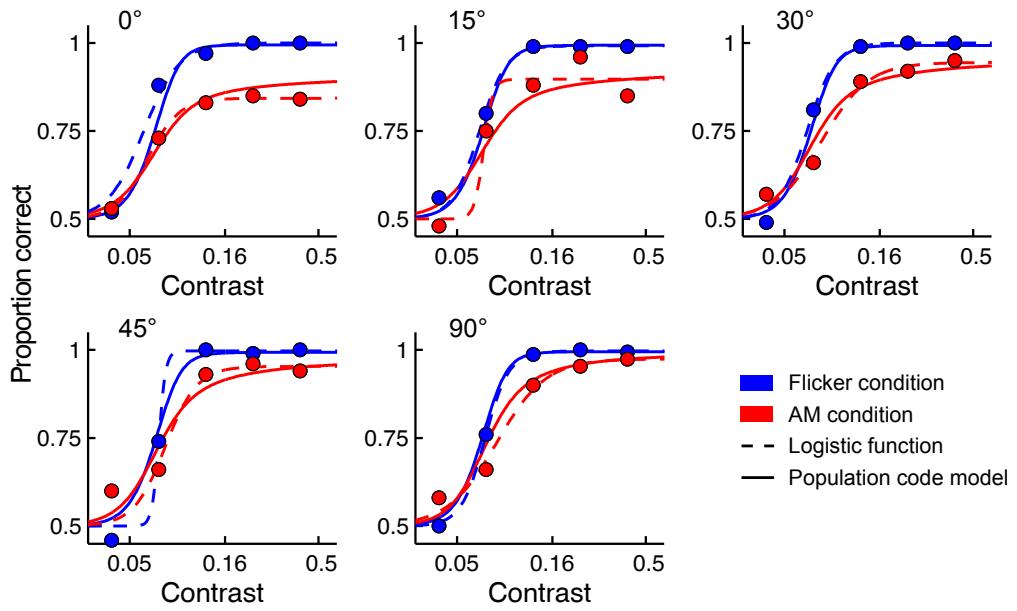


Figure 4.9. Model fits to data of observer AV. Plots are shown for each orientation difference between the target and inducer gratings. Red and blue symbols denote the AM and Flicker conditions respectively. Dashed lines depict the best-fitting logistic psychometric functions, while full lines represent the best-fitting population code model.

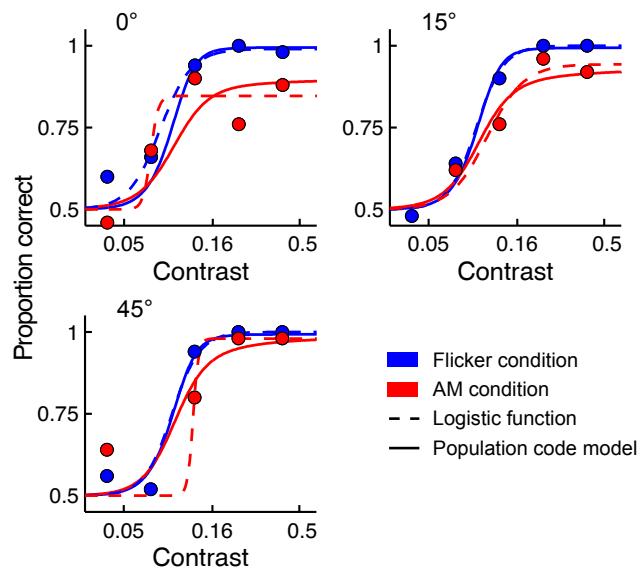


Figure 4.10. Model fits to data of observer CV. Plots are shown for each orientation difference between the target and inducer gratings (no data was collected for this observer for orientation differences of 30° and 90°). Red and blue symbols denote the AM and Flicker conditions respectively. Dashed lines depict the best-fitting logistic psychometric functions, while full lines represent the best-fitting population code model.

References

- Ahmed, B., Hanazawa, A., Undeman, C., Eriksson, D., Valentiniene, S., & Roland, P. E. (2008). Cortical dynamics subserving visual apparent motion. *Cerebral Cortex*, 18(12), 2796–2810.
- Albrecht, D. G., & Hamilton, D. B. (1982). Striate cortex of monkey and cat: contrast response function. *Journal of Neurophysiology*, 48(1), 217–237.
- Alink, A., Schwiedrzik, C. M., Kohler, A., Singer, W., & Muckli, L. (2010). Stimulus predictability reduces responses in primary visual cortex. *Journal of Neuroscience*, 30(8), 2960–2966.
- Bonds, A. (1989). Role of inhibition in the specification of orientation selectivity of cells in the cat striate cortex. *Visual Neuroscience*, 2(1), 41–55.
- Carandini, M., Heeger, D., & Movshon, J. (1997). Linearity and normalization in simple cells of the macaque primary visual cortex. *Journal of Neuroscience*, 17(21), 8621–8644.
- Chaumon, M., & Busch, N. A. (2014). Prestimulus neural oscillations inhibit visual perception via modulation of response gain. *Journal of Cognitive Neuroscience*, 26(11), 2514–2529.
- Chirimuuta, M., & Tolhurst, D. J. (2005). Does a Bayesian model of V1 contrast coding offer a neurophysiological account of human contrast discrimination? *Vision Research*, 45(23), 2943–2959.
- Cohen, M., & Newsome, W. (2009). Estimates of the contribution of single neurons to perception depend on timescale and noise correlation. *Journal of Neuroscience*, 29(20), 6635–6648.
- De Valois, R. L., Albrecht, D. G., & Thorell, L. G. (1982). Spatial frequency selectivity of cells in macaque visual cortex. *Vision Research*, 22(5), 545–559.
- De Valois, R. L., & De Valois, K. K. (1988). *Spatial vision*. Oxford University Press, USA.
- De Valois, R. L., Yund, E. W., & Hepler, N. (1982). The orientation and direction selectivity of cells in macaque visual cortex. *Vision Research*, 22(5), 531–544.
- Fang, F., Kersten, D., & Murray, S. O. (2008). Perceptual grouping and inverse fMRI activity patterns in human visual cortex. *Journal of Vision*, 8(7), 2–9.
- Foley, J. M. (1994). Human luminance pattern-vision mechanisms: masking experiments require a new model. *Journal of the Optical Society of America*, 11(6), 1710–1719.
- Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1456), 815–836.
- Geisler, W. S., & Albrecht, D. G. (1997). Visual cortex neurons in monkeys and cats: detection, discrimination, and identification. *Visual Neuroscience*, 14(5), 897–919.
- Goebel, R., Khorram-Sefat, D., Muckli, L., Hacker, H., & Singer, W. (1998). The constructive nature of vision: direct evidence from functional magnetic resonance imaging studies of apparent motion and motion imagery. *European Journal of Neuroscience*, 10(5), 1563–1573.
- Goris, R. L., Putzeys, T., Wagemans, J., & Wichmann, F. A. (2013). A neural population model for visual pattern detection. *Psychological Review*, 120(3), 472–496.

- Goris, R. L., Wichmann, F., & Henning, G. (2009). A neurophysiologically plausible population code model for human contrast discrimination. *Journal of Vision*, 9(7), 1–22.
- Harrison, L., Stephan, K., Rees, G., & Friston, K. (2007). Extra-classical receptive field effects measured in striate cortex with fMRI. *Neuroimage*, 34(3), 1199–1208.
- Heeger, D. (1992). Half-squaring in responses of cat striate cells. *Visual Neuroscience*, 9(5), 427–443.
- Herrmann, K., Montaser-Kouhsari, L., Carrasco, M., & Heeger, D. J. (2010). When size matters: attention affects performance by contrast or response gain. *Nature Neuroscience*, 13(12), 1554–1559.
- Hidaka, S., Nagai, M., Sekuler, A. B., Bennett, P. J., & Gyoba, J. (2011). Inhibition of target detection in apparent motion trajectory. *Journal of Vision*, 11(10), 1–12.
- Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *Journal of Physiology*, 148(3), 574–591.
- Kanai, R., & Kamitani, Y. (2003). Time-locked perceptual fading induced by visual transients. *Journal of Cognitive Neuroscience*, 15(5), 664–672.
- Katzner, S., Busse, L., & Carandini, M. (2011). GABA-A inhibition controls response gain in visual cortex. *Journal of Neuroscience*, 31(16), 5931–5941.
- Larsen, A., Madsen, K., Ellegaard Lund, T., & Bundesen, C. (2006). Images of illusory motion in primary visual cortex. *Journal of Cognitive Neuroscience*, 18(7), 1174–1180.
- Liu, T., Slotnick, S. D., & Yantis, S. (2004). Human MT+ mediates perceptual filling-in during apparent motion. *Neuroimage*, 21(4), 1772–1780.
- Matthews, W. J., Terhune, D. B., van Rijn, H., Eagleman, D. M., Sommer, M. A., & Meck, W. H. (2014). Subjective duration as a signature of coding efficiency: emerging links among stimulus repetition, predictive coding, and cortical GABA levels. *Timing & Time Perception Reviews*, 1(5), 1–12.
- Michel, M., & Geisler, W. S. (2011). Intrinsic position uncertainty explains detection and localization performance in peripheral vision. *Journal of Vision*, 11(1), 1–18.
- Mikami, A., Newsome, W. T., & Wurtz, R. H. (1986). Motion selectivity in macaque visual cortex. II. Spatiotemporal range of directional interactions in MT and V1. *Journal of Neurophysiology*, 55(6), 1328–1339.
- Muckli, L., Kohler, A., Kriegeskorte, N., & Singer, W. (2005). Primary visual cortex activity along the apparent-motion trace reflects illusory perception. *PLoS Biology*, 3(8), 1501–1510.
- Muckli, L., Kriegeskorte, N., Lanfermann, H., Zanella, F. E., Singer, W., & Goebel, R. (2002). Apparent motion: event-related functional magnetic resonance imaging of perceptual switches and states. *Journal of Neuroscience*, 22(9), RC219.
- Murray, S. O., Kersten, D., Olshausen, B. A., Schrater, P., & Woods, D. L. (2002). Shape perception reduces activity in human primary visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 99(23), 15164–15169.
- Nachmias, J., & Sansbury, R. V. (1974). Grating contrast: discrimination may be better than detection. *Vision Research*, 14(10), 1039–1042.
- Najemnik, J., & Geisler, W. S. (2005). Optimal eye movement strategies in visual search. *Nature*, 434, 387–391.

- Naka, K., & Rushton, W. (1966). S-potentials from colour units in the retina of fish (Cyprinidae). *Journal of Physiology*, 185(3), 536–555.
- Pelli, D. G. (1985). Uncertainty explains many aspects of visual contrast detection and discrimination. *Journal of the Optical Society of America A, Optics and Image Science*, 2(9), 1508–1532.
- Putzeys, T., Bethge, M., Wichmann, F., Wagemans, J., & Goris, R. (2012). A new perceptual bias reveals suboptimal population decoding of sensory responses. *PLoS Computational Biology*, 8(4), e1002453.
- Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2(1), 79–87.
- Rosenberg, A., Patterson, J. S., & Angelaki, D. E. (2015). A computational perspective on autism. *Proceedings of the National Academy of Sciences of the United States of America*, 112(30), 9158–9165.
- Schellekens, W., van Wezel, R. J., Petridou, N., Ramsey, N. F., & Raemaekers, M. (2014). Predictive coding for motion stimuli in human early visual cortex. *Brain Structure and Function*, 1–12.
- Swindale, N. V. (1998). Orientation tuning curves: empirical description and estimation of parameters. *Biological Cybernetics*, 78(1), 45–56.
- Vetter, P., Grosbras, M.-H., & Muckli, L. (2015). TMS over V5 disrupts motion prediction. *Cerebral Cortex*, 25(4), 1052–1059.
- Vogels, R., Spileers, W., & Orban, G. A. (1989). The response variability of striate cortical neurons in the behaving monkey. *Experimental Brain Research*, 77(2), 432–436.
- Watson, A. B., & Solomon, J. A. (1997). Model of visual contrast gain control and pattern masking. *Journal of the Optical Society of America A, Optics and Image Science*, 14(9), 2379–2391.
- Webb, B. S., Tinsley, C. J., Barracough, N. E., Parker, A., & Derrington, A. M. (2003). Gain control from beyond the classical receptive field in primate primary visual cortex. *Visual Neuroscience*, 20(3), 221–230.
- Wertheimer, M. (1912). Experimentelle Studien über das Sehen von Bewegung. *Zeitschrift für Psychologie*, 61, 161–265.
- Wibral, M., Bledowski, C., Kohler, A., Singer, W., & Muckli, L. (2009). The timing of feedback to early visual cortex in the perception of long-range apparent motion. *Cerebral Cortex*, 19(7), 1567–1582.
- Wichmann, F. A., & Hill, N. J. (2001a). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception and Psychophysics*, 63(8), 1293–1313.
- Wichmann, F. A., & Hill, N. J. (2001b). The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Perception and Psychophysics*, 63(8), 1314–1329.
- Williford, T., & Maunsell, J. H. (2006). Effects of spatial attention on contrast response functions in macaque area V4. *Journal of Neurophysiology*, 96(1), 40–54.
- Yantis, S., & Nakama, T. (1998). Visual interactions in the path of apparent motion. *Nature Neuroscience*, 1(6), 508–512.
- Zohary, E., Shadlen, M. N., & Newsome, W. T. (1994). Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature*, 370, 140–143.

Chapter 5

Response suppression during apparent motion reduces perceived grating contrast

This chapter is a manuscript in preparation:

Van Humbeeck, N., Putzeys, T. & Wagemans, J. (2016). *Response suppression during apparent motion reduces perceived grating contrast*. Manuscript in preparation.

5.1 Introduction

In Chapter 4, we investigated the phenomenon of apparent motion (AM), which entails the illusory perception of a moving visual stimulus occurring when two stationary stimuli are alternately presented at different spatial locations. Observers report perceiving continuous motion along a linear path connecting the two AM-inducing stimuli. Behavioral as well as physiological studies have attributed AM to excitation of neurons in the early visual system (Hidaka, Nagai, Sekuler, Bennett, & Gyoba, 2011; Souto & Johnston, 2012; Yantis & Nakama, 1998). More specifically, it has been claimed that the AM-inducing stimuli cause a response of neurons in primary visual cortex (V1) that are tuned to spatial locations along the AM path. We have challenged this claim in Chapter 4, reporting behavioral data and computational modelling results suggesting that AM involves significant suppression of responses in the early visual system instead of excitation.

In the present study, we start from the population code model of Chapter 4 and formulate a new prediction concerning the effect of AM on the apparent or perceived contrast of a grating. Multiple studies have investigated the perceived contrast of simple grating stimuli in the context of surround masking. When a grating stimulus is surrounded by another, high-contrast grating pattern, its perceived contrast is reduced significantly (Cannon & Fullenkamp, 1991; Chubb, Sperling, & Solomon, 1989; Snowden & Hammett, 1998; Solomon, Sperling, & Chubb, 1993; Xing & Heeger, 2000; Zenger-Landolt & Heeger, 2003). This perceived contrast reduction has been attributed to response suppression in V1 caused by the surrounding grating pattern. It appears that suppression of V1 responses to a grating lowers its apparent contrast. As we found in Chapter 4 that presenting a grating on the AM path suppresses the V1 responses to that grating, we expect to find a reduction of apparent contrast in the presence of AM. This prediction was tested in the present study. Observers performed a grating contrast discrimination task in the presence of AM. As in Chapter 4, an AM path was created on both sides of the screen. Two gratings were presented simultaneously in the middle of the two paths. Importantly, the orientation of one of the gratings matched the orientation of the AM inducers, while the orientation of the other grating was orthogonal to the inducer orientation.

Given the results of Chapter 4, we expect V1 responses to the gratings to be suppressed significantly, but only when its orientation matches the inducer orientation. Consequently, we predict that the perceived contrast of the grating will be lower when its orientation matches that of the inducers. This is exactly what we found in the present study.

5.2 Materials and Methods

5.2.1 Subjects

Seven observers (AW, BO, JM, LM, ML, TV, VG) participated in the experiment. All had normal or corrected-to-normal vision and were naive to the design and the purpose of the experiment (two male and five female, age range 20-23). All participants gave their written consent before the start of the experiment. Observers were paid 8 euros an hour for participating. The experiment began with a block of 50 practice trials in order for participants to become familiar with the stimuli and task. All participants reported having a clear AM percept.

5.2.2 Apparatus

Hardware and software were identical to those used in Chapter 4. Participants sat in a darkened room with their heads stabilized using a chin rest, at a viewing distance of 60 cm (corresponding to a pixel size of 0.0315° of visual angle). The mean background luminance of the screen was equal to 72.5 cd/m^2 . A Cedrus response box (RB-530, Cambridge Research Systems) registered participants' responses.

5.2.3 Stimuli

The stimuli in this study consisted of Gabor patches, which were the product of a cosine grating and a 2D Gaussian envelope ($\text{SD} = 0.75^\circ$). The gratings had a spatial frequency of 1.5 cycles per degree. Stimuli were displayed on a gray background (Michelson con-

trast of 50%). A reference grating was presented at one side from a fixation cross ($0.76^\circ \times 0.76^\circ$) at 10° eccentricity, while a comparison grating was simultaneously presented at the other side. Reference and comparison gratings were assigned to both sides randomly across trials. The orientation of the reference and comparison grating always equalled 0° (horizontal). Both these gratings were positioned exactly in between two AM-inducing stimuli, which were vertically separated by 8° . The AM-inducing stimuli positioned above and below the reference grating were oriented horizontally, thus matching the orientation of the reference grating. The AM-inducing stimuli located above and below the comparison grating were oriented orthogonally to the comparison grating. In a control condition, the reference and comparison gratings were presented in between two Flicker-inducing stimuli (see Chapter 4 and below). The inducing stimuli had a Michelson contrast of 100%. The reference grating always had a Michelson contrast of 15%, while the contrast levels of the comparison grating ranged from 5% to 50% Michelson contrast.

5.2.4 Discrimination task

In a spatial two-alternative forced-choice (2AFC) task, observers were asked to indicate which of two gratings, i.e., the reference or comparison grating, had the highest contrast. A trial started with a fixation cross that was shown for 500 ms. The inducer stimuli were then presented for a duration of 80 ms alternately at the top and bottom position, each four times with an inter-stimulus interval of 106 ms. This resulted in the percept of a vertically moving grating along the path between the inducers. The reference and comparison gratings were flashed briefly for 30.8 ms during the fourth sequence of the inducers, 38 ms after the presentation of the inducer positioned at the top and at an intermediate temporal position in the interstimulus interval between the two inducers. At the end of the trial, observers were asked to indicate which side of the screen contained the grating with the highest contrast. Observers received auditory feedback after each trial. A Flicker control condition was included, for which the procedure was identical, except that the inducers were simultaneously shown at the top and bottom position (see also Chapter 4). This resulted in a complete disruption of AM. Figure 5.1 shows an illustration of the stimulus sequence in the AM and Flicker condition.

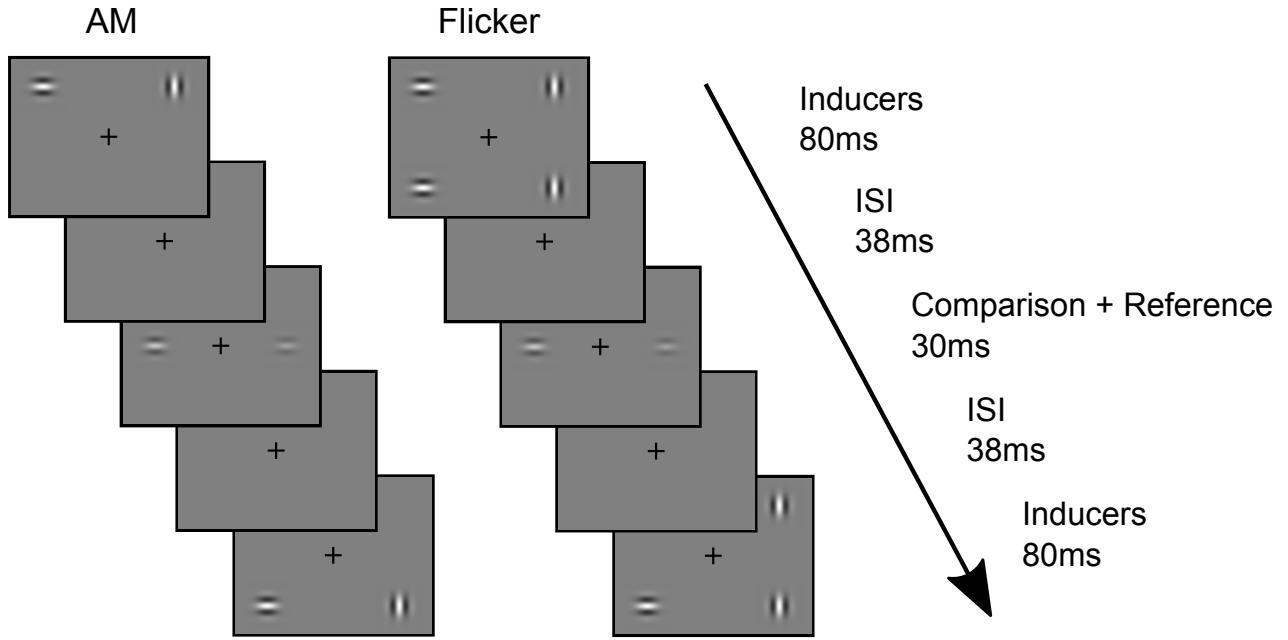


Figure 5.1. Illustration of the stimulus sequence in the AM and Flicker condition of the 2AFC discrimination task. Contrast of the reference and comparison grating has been increased for illustration purposes. A trial consisted of four of these stimulus sequences. The reference and comparison gratings were only presented during the fourth stimulus sequence.

The experiment consisted of blocks of 50 trials, in which the physical contrast of the comparison grating (6 contrast levels) and condition (AM or Flicker) were randomised. Each observer completed 200 trials for each combination of condition and contrast level, yielding a total of 2400 trials for each observer.

5.2.5 Population code model

Equations

We used a similar population code model as the one used in Chapter 4. The encoding stages of both models are identical. With respect to the decoding stage, recall from Chapter 4 that a simple linear decoder was used which considers the two spatial locations at which a single target grating could appear. In the current study, observers performed a discrimination task instead of a detection task. Consequently, a grating is presented simultaneously at both spatial locations. The task of the decoder is now to select the location containing the grating of the highest contrast. To accomplish this, a linear

decoding strategy was implemented that is highly similar to the one used for detection. The decoder sums the responses of two subpopulations of neurons, each tuned to one of the two grating locations. The location yielding the largest summed response is indicated as containing the highest-contrast stimulus. Previous studies have used a similar decoder in a population code model to account for contrast discrimination performance (Goris, Wichmann, & Henning, 2009).

S_{comp} is defined as the summed response of those neurons tuned to the location at which the comparison grating is presented, where S_{ref} equals the summed response of neurons tuned to the reference grating location. Similar to Equations 4.11 and 4.12 in Chapter 4, the average values of these variables in the AM condition are provided by:

$$\overline{S_{comp}}(c, \theta_{comp}, \theta_{ind,comp}) = \sum_{i=1}^N \bar{r}_i(c, \theta_{comp}, \theta_{ind,comp}) \quad (5.1)$$

$$\overline{S_{ref}}(c, \theta_{ref}, \theta_{ind,ref}) = \sum_{i=1}^N \bar{r}_i(c, \theta_{ref}, \theta_{ind,ref}) \quad (5.2)$$

where θ_{comp} and θ_{tgt} respectively denote the orientation of the comparison and reference grating. Both parameters equal 0° in our experiment (horizontal orientation). $\theta_{ind,comp}$ and $\theta_{ind,ref}$ represent the orientation of the inducer gratings surrounding respectively the comparison and reference grating. The value of these parameters equals 90° (vertical orientation) and 0° , respectively. The average response \bar{r}_i is again provided by Equation 4.10. As in Chapter 4, the standard contrast response function of Equation 4.7 was used to obtain $\overline{S_{comp}}$ and $\overline{S_{ref}}$ for the Flicker condition. Note that the function $\overline{S_{ref}}$ is always evaluated for $c = 0.15$, i.e., the constant value of the reference grating, except when generating plots showing $\overline{S_{ref}}$ for a range of different contrast values.

Following Vogels, Spileers, and Orban (1989), the variance of the summed responses is given by:

$$var(S_{comp}) = \zeta \overline{S_{comp}} \quad (5.3)$$

and

$$var(S_{ref}) = \zeta \overline{S_{ref}} \quad (5.4)$$

The predicted proportion p of trials on which the comparison grating will be chosen as the grating of the highest contrast is then provided by the cumulative Gaussian function:

$$p = \int_0^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \quad (5.5)$$

with

$$\mu = \overline{S_{comp}} - \overline{S_{ref}} \quad (5.6)$$

and

$$\sigma = \frac{1}{\epsilon} \sqrt{\text{var}(S_{comp}) + \text{var}(S_{ref})} \quad (5.7)$$

where ϵ is an efficiency parameter (see Chapter 4).

Our contrast discrimination decoder only takes a-priori knowledge of the grating location into account and disregards the neurons' orientation tuning. It sums the responses of all neurons tuned to either the comparison or the reference grating location, irrespective of whether the neurons are sensitive to the orientation of those gratings. A similar approach was followed to model the contrast detection experiment in Chapter 4. However, the target grating orientation was randomized across trials. It would be impossible in that case for any decoder to select neurons sensitive to the target orientation without first detecting the target. In the present contrast discrimination experiment, the orientation of the comparison and reference grating is constant throughout the experiment. A linear decoder may thus be able to preferentially weigh the responses of optimally-tuned neurons when computing S_{comp} and S_{ref} . Previous studies have shown, however, that human observers fail to use readily available knowledge of low-level stimulus properties when decoding V1 responses in simple detection and discrimination tasks (Putzeys, Bethge, Wichmann, Wagemans, & Goris, 2012). An earlier population code model of contrast discrimination also incorporates a simple decoder that is agnostic about grating orientation when integrating V1 responses (Goris et al., 2009).

It should be noted that we assume the summed responses S_{comp} and S_{ref} to be directly proportional to the perceived contrast of respectively the comparison and the reference grating. Haun and Peli (2013) made a similar assumption in their model of perceived

contrast, although these authors also considered the possibility of non-linear summation of V1 responses.

Deriving relative matching contrast

We use the model prediction to derive the so-called relative matching contrast. This is the comparison grating contrast that is required to reach a comparison choice proportion of 50%. When the comparison grating is presented at the relative matching contrast, the perceived contrast of the comparison grating is equal to that of the reference grating as both gratings are selected with a probability of 50%. The relative matching contrast can be interpreted as the perceived contrast of the reference grating relative to that of the comparison grating. To obtain the relative matching contrast, we linearly interpolated the predicted comparison choice proportions computed for a wide range of comparison grating contrasts.

Model constraints and fitting

The constraints listed in Chapter 4 were also adopted in the present study. The parameters k_{inh} , r_{max} , ζ were all fixed to the physiologically plausible values reported in Chapter 4. However, some additional parameters were fixed as the current data did not provide sufficient constraints. Changing any of these parameters does not affect the conclusions of the present study. The concentration parameter k_{exc} controlling the orientation bandwidth of the excitatory receptive fields was set to 1.35, while the spontaneous discharge rate r_0 was set to 4.55%. Both values were estimated in Chapter 4 and are consistent with physiological observations in V1. Our model did not require AM-induced inhibition, i.e., the parameter β could be constrained to zero without reducing the quality of the model fit. We therefore omitted AM-induced inhibition from the model as eliminating a free parameter reduces model complexity and, consequently, minimizes the risk of overfitting.

In total, 6 free parameters were estimated using a maximum-likelihood fitting procedure (Wichmann & Hill, 2001a). We performed multiple fits using randomized starting values for each parameter. The deviance statistic D (Wichmann & Hill, 2001a) was used to

quantify the goodness-of-fit of a model. When provided, the average D value refers to the D value averaged across subjects.

A parametric Monte-Carlo bootstrap procedure involving 1000 samples provided the deviance statistic distribution (Wichmann & Hill, 2001b). This bootstrapped distribution represents the range of deviance values to be expected if the model under consideration actually underlies the observed data. For each model fit, we report the p value corresponding to the proportion of bootstrapped deviance values that are equal to or higher than the model fit deviance D . A model fit is deemed acceptable when p is larger than or equal to 0.05. Note that Bonferroni-correction is applied whenever statistical tests are performed simultaneously for all subjects. The parametric Monte-Carlo bootstrap procedure was also used to obtain confidence intervals for estimates of the parameters as well as the relative matching contrasts.

Akaike's Information Criterion (AIC) was used to compare the goodness-of-fit of two models while taking into account the difference in free parameters (which is an indication of model complexity). More specifically, we compared the AIC of the model of the current study and the original model of Chapter 4, which used 10 free parameters and included the AM-induced inhibition effect.

5.3 Results

5.3.1 AM biases choice proportions

Figure 5.2 shows the data of the discrimination task. Results were highly consistent for 5 out of 7 observers. The results of these 5 observers are discussed in the subsequent sections, while the data of the two remaining subjects are detailed in Appendix A (section 5.5). The symbols in Figure 5.2 show the comparison choice proportion, i.e., the proportion of trials on which the comparison grating was indicated to have the highest contrast, as a function of the actual comparison grating contrast. For both the AM condition and the Flicker control condition, comparison choice proportion increased with comparison contrast. Observers increasingly chose the comparison grating as the highest-contrast

grating as its contrast was increased. In the AM condition, however, comparison choice proportion was higher at low comparison contrasts. Choice proportion never dropped below 25% for the majority of the observers. It appears that observers were biased to choose the comparison grating over the reference grating, even though the comparison contrast was significantly lower than the reference contrast of 15%.

In the AM condition, both reference and comparison gratings were presented in the presence of AM. The comparison grating orientation, however, was orthogonal to the inducer grating orientation while the reference grating orientation matched that of the inducers. As detailed in the introduction, we expect the perceived contrast of the reference grating to be reduced because of response suppression introduced by AM. Because of the orientation difference between the comparison and inducer gratings, the perceived contrast of the comparison grating should be less affected. In other words, we expect the perceived contrast of the comparison grating to be higher than that of the reference grating. Such a difference in perceived contrast may well explain the high comparison choice proportions. Arguably, AM reduces the perceived contrast of the reference grating relative to the comparison grating, causing observers to choose the latter stimulus as the highest-contrast grating.

5.3.2 AM reduces perceived grating contrast

To confirm our intuition mentioned in the previous section, we fitted the population code model to the data. The model fit is acceptable (average $D = 11.8$, $p > 0.05$ for all observers). Figure 5.2 (full lines) show the model predictions. The model accurately captures the increase in comparison choice proportion with increasing comparison contrast, as well as the difference in choice proportions between the AM and Flicker condition. The model predicts the bias in choice proportions, i.e., the increased comparison choice proportions in the AM condition at low comparison contrasts.

Figure 5.3 shows the relative matching contrast predicted by the model. The estimated values are significantly lower than 15% ($p < 0.05$ for all observers), which is the actual contrast of the reference grating. This indicates that the physical contrast of the compari-

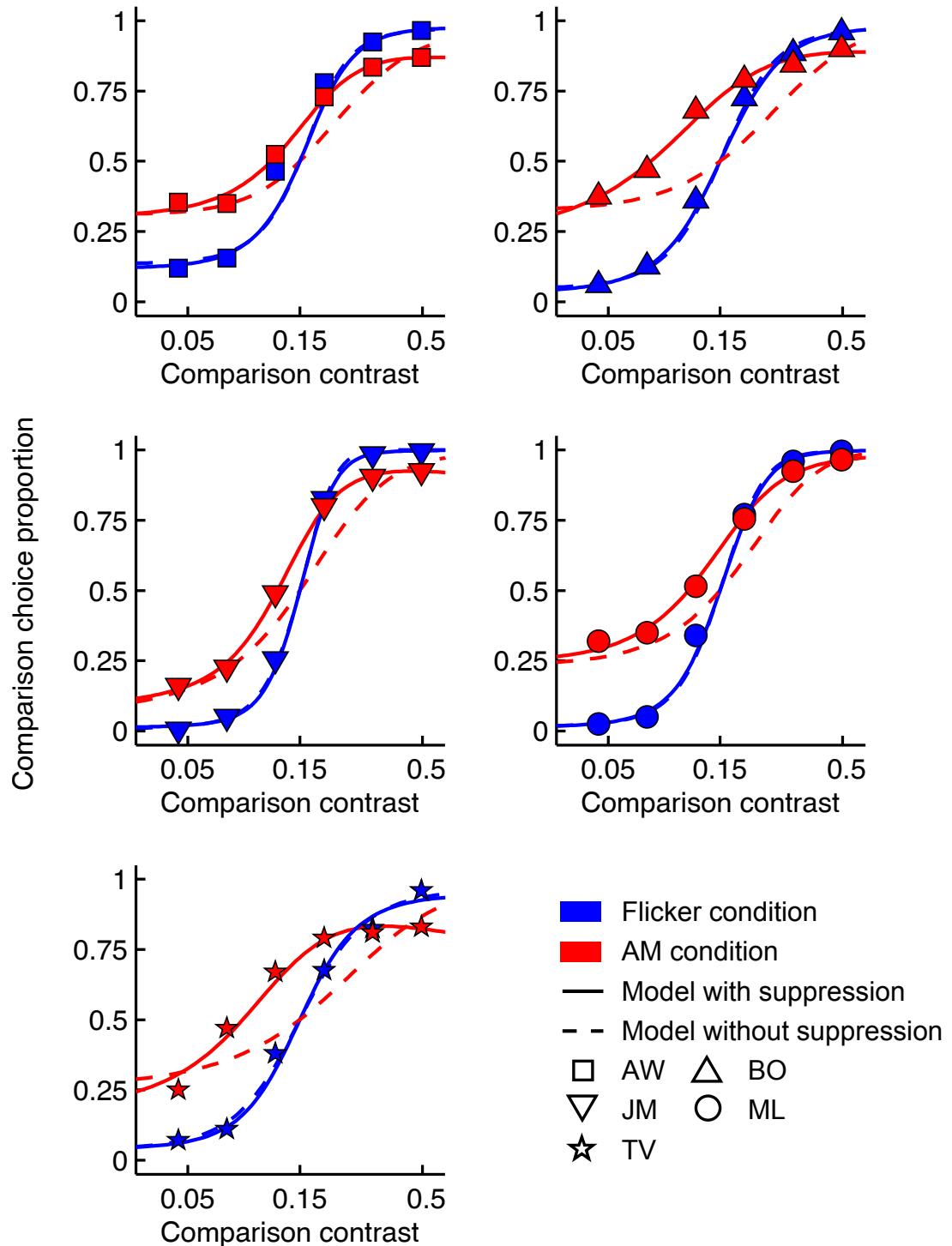


Figure 5.2. Individual data of five observers and model fits. Red and blue lines and symbols represent the AM and Flicker conditions, respectively. Full lines depict the predicted comparison choice proportions for our population code model incorporating AM-induced suppression, while dashed lines denote the predicted comparison choice proportions for a population code model that does not include AM-induced suppression ($\gamma = 0$).

son grating should be lower than the physical contrast of the reference grating in order for it to match the perceived contrast of the reference grating. In other words, the perceived contrast of the reference grating contrast is lower than that of the comparison grating when both gratings are presented at the same 15% contrast level. This observation is consistent with the claim that AM lowers the perceived contrast of a grating (i.e., the reference grating) presented on the AM path when its orientation matches the inducer grating orientation. When the orientation of the same grating presented on the AM path is orthogonal to the inducer orientation, its perceived contrast is less affected. This leads to a difference in perceived contrast of the two gratings, and consequently, to a relative matching contrast deviating from the actual reference grating contrast.

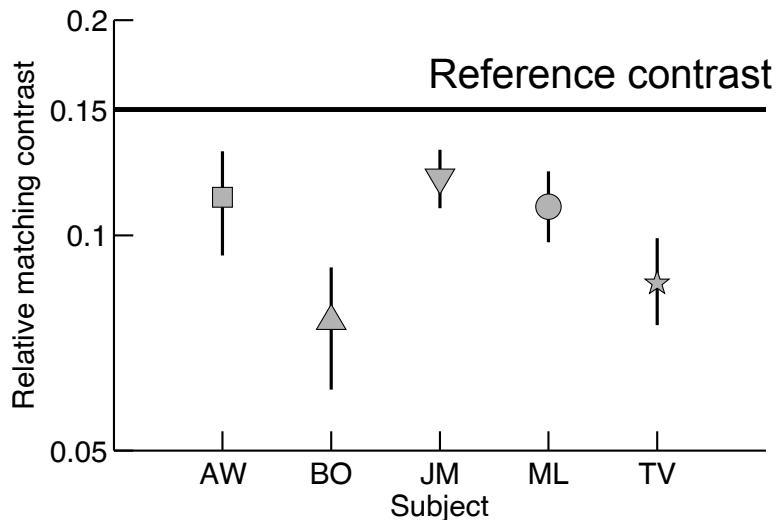


Figure 5.3. Relative matching contrast predicted by the population code model. The horizontal line denotes the reference grating contrast. For each observer, the physical contrast of the comparison grating should be significantly lower than the physical contrast of the reference grating in order for it to match the perceived contrast of the reference grating. Error bars denote the 95% confidence intervals (Bonferroni corrected).

5.3.3 Perceived grating contrast reduction is caused by response suppression

To understand how AM affects perceived contrast, we analysed the V1 contrast responses predicted by the model. The response exponent p equals 2.40 on average across subjects ($SEM = 0.174$). Estimates for each subject are shown in Figure 5.4. The semi-saturation

contrast c_{50} is equal to 19.59% ($SEM = 0.8\%$). These values are well within the range of values reported in physiological single-cell recording studies measuring actual monkey V1 contrast response functions (Chirimutta & Tolhurst, 2005; Heeger, 1992).

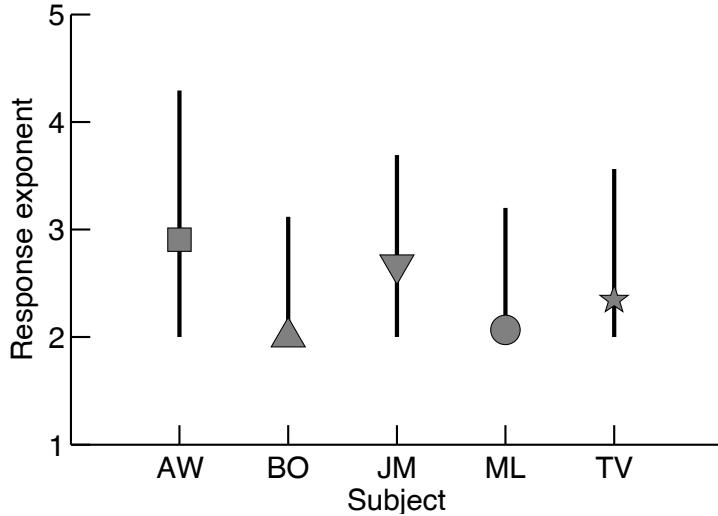


Figure 5.4. Estimates of the response exponent in our population code model for each observer. Error bars denote the 95% confidence intervals (Bonferroni corrected).

According to our model, choice proportions in the discrimination task ultimately depend on S_{comp} and S_{ref} (see Equations 5.1 and 5.2), i.e., the contrast response to respectively the comparison and the reference grating, summed across all neurons tuned to the respective grating locations. Figure 5.5A shows the average value of S_{comp} (green full line) and S_{ref} (purple full line) as a function of contrast in the AM condition for subject ML. In the Flicker condition, S_{comp} and S_{ref} are by definition equal (grey full line). A first observation is that the summed response functions in the AM condition are scaled down versions of the function in the Flicker condition. This is the result of AM-induced response gain suppression. The V1 contrast response to a grating is suppressed when this grating is presented on the AM path. The level of suppression is controlled by the γ parameter of the model's encoding stage (see Chapter 4). This parameter equals 76.52% on average ($SEM = 1.08\%$), which means that the contrast response is scaled down by a factor of $(100 - 76.52)\% = 23.48\%$. For subject ML, γ is estimated at 74.89%. Figure 5.6 shows the estimates of γ for all subjects. Note that as the contrast responses of individual neurons are suppressed, the summed response is reduced as well.

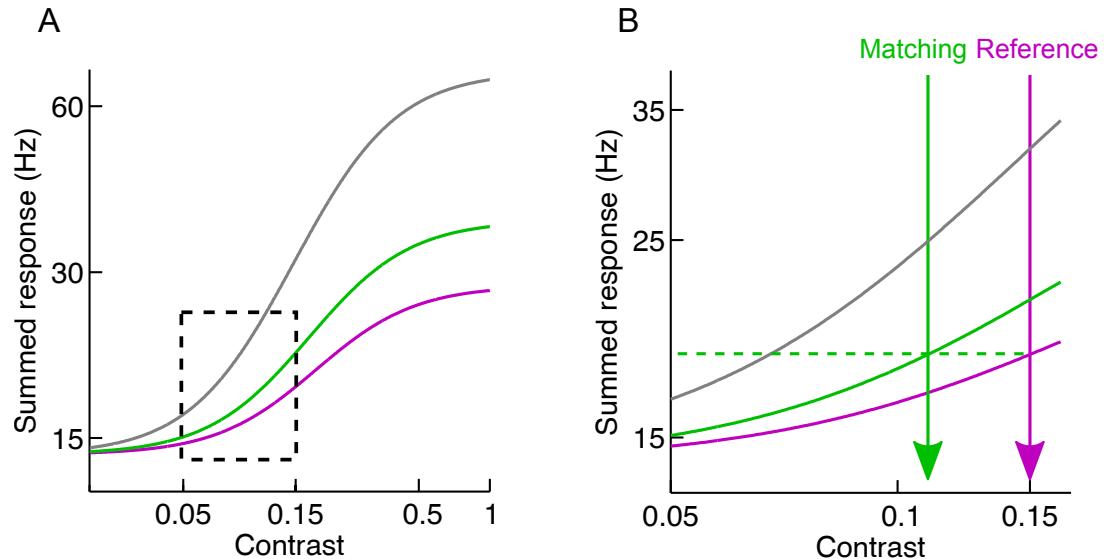


Figure 5.5. (A) Average summed contrast response to the comparison and the reference grating tuned to the respective grating locations (i.e., S_{comp} and S_{ref}) as a function of contrast for subject ML. The green full line depicts the average value of S_{comp} and the purple full line shows the average value of S_{ref} as a function of contrast in the AM condition. The grey full line depicts the average value of S_{comp} and S_{ref} in the Flicker condition, which are identical according to the model. (B) A detailed view of part of the summed contrast response functions for subject ML (indicated by the dashed rectangle in panel A). The green vertical line indicates at which contrast level the comparison grating has to be presented to obtain a summed response equal to that of the reference grating (which has a contrast level of 15%). 11% is the relative matching contrast: the perceived contrast of the comparison grating with a physical contrast of 11% will be equal (horizontal green line) to that of the reference grating with a physical contrast of 15%.

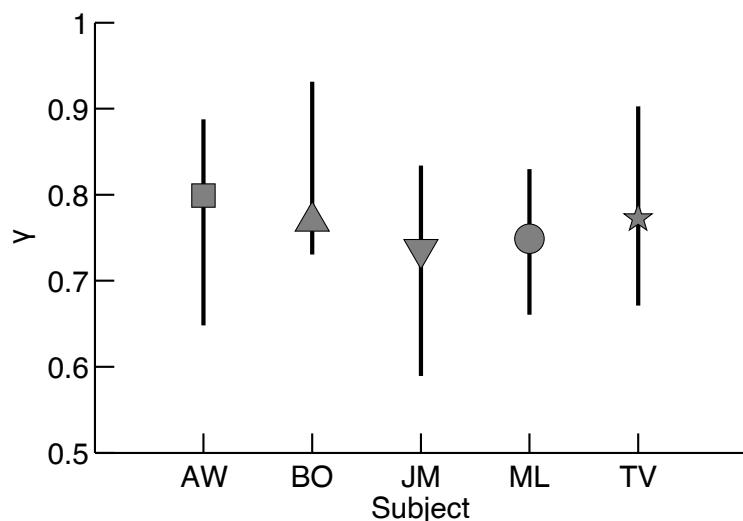


Figure 5.6. Estimates of AM-induced suppression (γ) in our population code model for each observer. Error bars denote the 95% confidence intervals (Bonferroni corrected).

In our model, AM-induced response gain suppression is tuned for orientation: the suppression is larger in neurons that are tuned to the orientation of the inducer gratings compared to neurons that are insensitive to the inducer orientation. The inducers surrounding the comparison grating have a vertical orientation. These vertical inducers cause strong suppression of neurons tuned to vertical orientations. However, neurons sensitive to horizontal orientations will be less suppressed. These latter neurons respond to the comparison grating, which is horizontally oriented. Consequently, the response to the comparison grating will be less suppressed than the response to the reference grating. Indeed, the reference grating is oriented horizontally and surrounded by horizontal inducers. The neurons responding to the reference grating will therefore be suppressed strongly. This explains why the summed response function S_{ref} for the reference grating is scaled down more than the function S_{comp} for the comparison grating. Due to the difference in inducer orientation and the resulting difference in the level AM-induced suppression, the neurons responding to the reference grating have a lower response gain than the neurons responding to the comparison grating. We return to the issue of orientation tuning in section 5.3.5.

The fact that the summed response function for the reference grating is reduced more than the comparison summed response function implies that the perceived contrast of the reference grating will be lower than that of the comparison grating when both gratings are presented at the same contrast. This explains why the relative matching contrast is lower than the actual reference grating contrast. To illustrate this point, Figure 5.5B provides a detailed view of part of the summed response functions of Figure 5.5A for subject ML. At a contrast level of 15% (the reference grating contrast, purple vertical line in Figure 5.5B), the average summed response to the reference grating equals 18.6 Hz, while the average summed response to the comparison grating is larger, equalling 21.4 Hz. To obtain a comparison grating summed response equal to that of the reference grating, the comparison grating contrast has to be reduced to 11% (green vertical line in Figure 5.5B). When the comparison grating is presented at this contrast level and the reference grating is presented at 15%, perceived contrast of both gratings will be equal and a comparison choice proportion of 50% is predicted. As such, the comparison

contrast of 11% is the relative matching contrast for subject ML.

Up until now, we have only discussed the contrast response function for subject ML. All other subjects also show the same pattern of scaled response functions, which determines the relative matching contrast in the way outlined above. However, two of these subjects do show a small but significant amount of AM-induced excitation that was not observed for subject ML. The effect of this excitation on perceived contrast is small compared to the suppressive effect and will be discussed in the next section.

The importance of AM-induced suppression in capturing perceived grating contrast and the relative matching contrast can be illustrated by fitting a version of the model in which suppression is removed ($\gamma = 0$). The dashed lines in Figure 5.2 show the predictions of this model. It can be seen that, in the absence of AM-induced suppression, the model consistently underestimates the comparison choice proportion in the AM condition and fails to predict a reduction in relative matching contrast. The reason is that perceived contrast of reference and comparison grating are predicted to be equal at the reference contrast level of 15%. This reduced model can be rejected for all subjects (average $D = 76.43$, $p < 0.001$ for all observers).

5.3.4 The role of AM-induced excitation

The level of AM-induced excitation is controlled by the parameter α of the model's encoding stage (see Chapter 4). The estimated values of this parameter are provided in Figure 5.7. Only observers JM and TV show a significant amount of excitation. The estimated α value equals 4.4% for observer JM and 8.8% for observer TV. Both values are significantly larger than zero ($p < 0.05$). For the remaining three observers, α does not differ significantly from zero. A version of the population code model in which $\alpha = 0$ could in principle be used for these observers. However, we refrained from doing so to be able to use the same model for all observers. Figure 5.8A shows the summed contrast response functions for subject JM (full lines). To illustrate the effect of excitation, we also show the model predictions obtained when AM-induced excitation was removed, i.e., when α was set to zero after fitting the model (dashed lines in Figure 5.8A). Figure 5.8B

shows the predicted comparison choice proportions obtained when $\alpha = 4.36\%$ (full lines) and $\alpha = 0$ (dashed lines). It can be seen from Figure 5.8B that removing excitation causes the model to overestimate comparison choice proportions at low and, to a lesser extent, at high comparison contrasts. The largest part of the predicted choice proportion function, i.e., the predictions for intermediate contrasts and the position of the function along the contrast axis (including the predicted relative matching contrast), remains relatively unaffected.

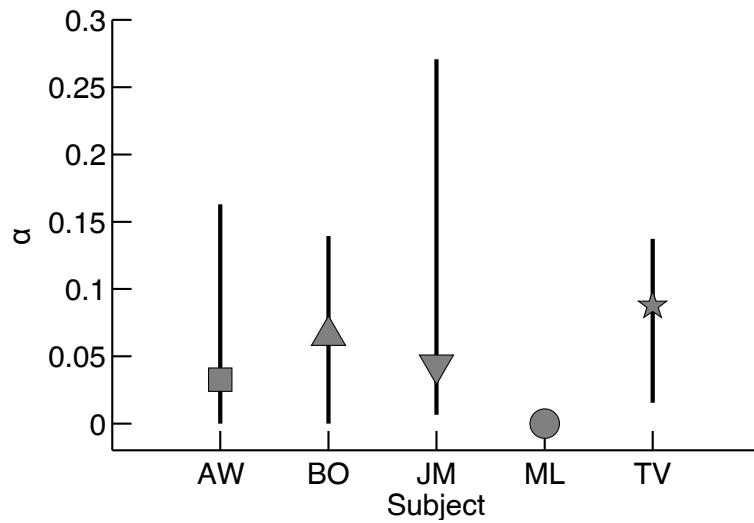


Figure 5.7. Estimates of AM-induced excitation (α) in our population code model for each observer. Error bars denote the 95% confidence intervals (Bonferroni corrected).

The overestimation of comparison choice proportions in the absence of excitation can be understood by considering the summed contrast response functions S_{comp} and S_{ref} , both in the presence and absence of excitation. For the reference grating, the contrast is always equal to 15%. This means that the summed response function S_{ref} has to be evaluated at a contrast of 15%. It is clear from Figure 5.8A that the effect of excitation on the S_{ref} function is relatively large at 15% contrast. The effect of excitation on the summed response function S_{comp} for the comparison grating is also large for intermediate comparison contrasts around 15%, but at low and high comparison contrasts, the response increase due to excitation is considerably smaller. The reason for this is the fact that V1 neurons' contrast response function is non-linear. Recall from Chapter 4 that, in the model's encoding stage, the excitation introduced by α is added to the actual grating contrast and then subjected to response expansion. Consequently, intermediate grating

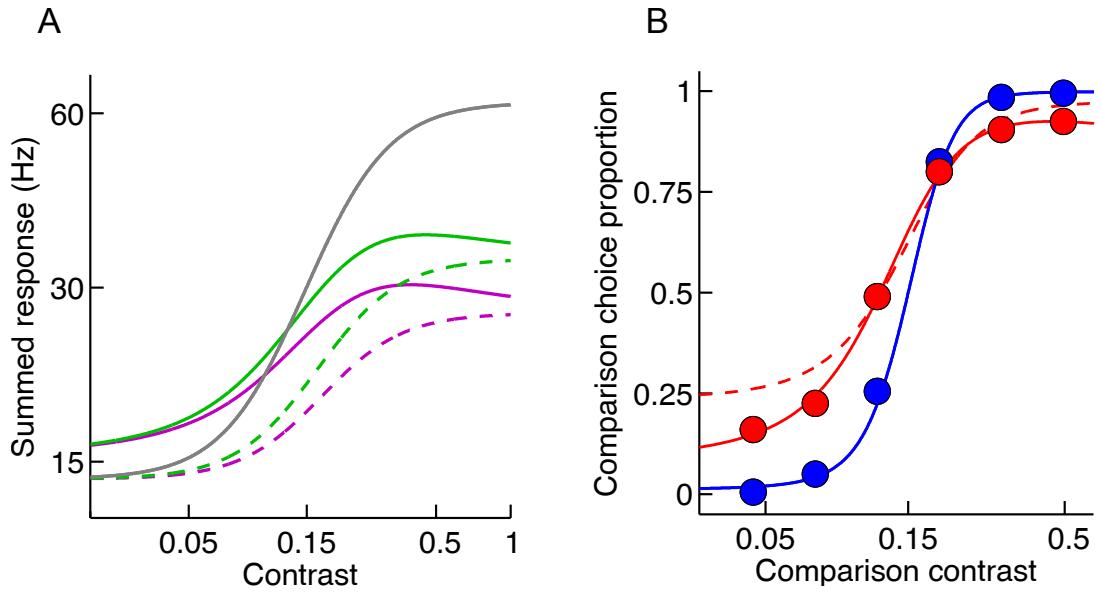


Figure 5.8. (A) Summed contrast response functions predicted by our population code model for subject JM when $\alpha = 4.36\%$ (full lines) and when α was set to zero after fitting the model (dashed lines). (B) Predicted comparison choice proportions obtained when $\alpha = 4.36\%$ and $\alpha = 0$ (depicted by full and dashed lines, respectively).

contrasts will act as a pedestal and amplify the excitation effect (e.g., Goris et al., 2009, see also section 4.3 of Chapter 4). Indeed, at intermediate grating contrasts, the model contrast response function operates in its steepest part. A small α -value will be converted into a large response increase in that case. At low and high contrasts, the contrast response function operates in its shallow parts and the same small α -value will have a lower impact on the contrast response.

In summary, at low and high comparison grating contrasts, excitation causes a strong increase in the summed response for the reference grating, while causing a smaller increase for the comparison grating. Consequently, AM-induced excitation increases the perceived contrast of the reference grating relative to that of the comparison grating, at least at low and high comparison contrasts. One should not conclude from this result that the net effect of AM is an increase of perceived grating contrast. As discussed in the previous section, strong AM-induced suppression causes a large reduction in the perceived contrast of both the reference and the comparison grating. The AM-induced excitation merely attenuates the suppressive effect at low and high comparison contrasts.

Note that the summed response functions for the reference and comparison gratings coin-

cide at low comparison contrasts, also in the presence of excitation. Given the orientation tuning of AM-induced excitation (see Chapter 4), one may expect excitation to be larger for the reference grating because of the match between the reference and inducer orientation. This intuition is incorrect: the orientation tuning of AM results in excitation of different subpopulations of neurons. For the reference grating, it is the subpopulation of neurons tuned to horizontal orientations that will be excited as the inducers are horizontally-oriented. In the case of the comparison grating, the vertical inducers will cause excitation in neurons tuned to vertical orientations. Ultimately, S_{comp} and S_{ref} are computed by summing the responses of all neurons tuned to the relevant grating locations. As such, it does not matter which subpopulation of neurons is excited. Excitation of either subpopulation will lead to the same increase in the summed response.

An interesting aspect of the summed contrast response functions in the AM condition is that they can decrease at high comparison contrasts. This is the case for the two subjects that show significant AM-induced excitation (see the green and purple full lines in Figure 5.8A for observer JM). The reason for the decrease is the inhibitory effect of contrast gain control. As mentioned in Chapter 4, the AM-induced excitation is tuned for orientation. However, as we will show in the next section, the tuning is relatively unspecific, i.e., AM induces excitation in a considerable amount of neurons that are insensitive to the inducer orientation. Like all other V1 neurons, these neurons are inhibited by the responses of a contrast gain control pool of neurons, which consist of neurons tuned to a wide range of orientations. Some of these inhibitory neurons will respond to the comparison grating. As such, as the comparison grating contrast is increased, the amount of inhibition in the gain control pool increases. This inhibition will (partially) counteract the AM-induced excitation, causing a reduction in the summed contrast response.

5.3.5 Orientation tuning of AM-induced suppression

As explained in section 5.3.3, the fact that the summed contrast response S_{ref} to the reference grating does not coincide with the summed response S_{comp} indicates that AM-suppression is tuned for orientation. The vertical inducers surrounding the comparison

grating will suppress neurons tuned to vertical orientations more than neurons tuned to other orientations. As such, the neurons tuned to the horizontal comparison grating orientation will be less suppressed compared to the neurons responding to the reference grating, which matches the orientation of its surrounding inducers.

An interesting question pertains to the extent to which the tuning is specific. Highly-specific tuning implies that AM-induced effects such as suppression and excitation are limited to a very narrow subpopulation of neurons that are exactly tuned to the inducer orientation. In the case of unspecific tuning, the subpopulation of neurons affected by AM may contain neurons tuned to a wider range of neurons. Vertical inducers may, for instance, suppress responses of neurons tuned to oblique or even horizontal orientations in that case. In Chapter 4, we found that the tuning function controlling the specificity of AM-induced excitation and suppression had a bandwidth equal to the orientation tuning bandwidth of a typical V1 excitatory receptive field. As V1 neurons are known to be tuned narrowly for orientation compared to neurons of other brain areas such as LGN and V2, the AM orientation tuning function uncovered in Chapter 4 can be considered highly selective. The value of the concentration parameter $k_{exc,AM}$ controlling the bandwidth was equal to 1.35 in the previous study. In the present study, the parameter estimate was considerably lower, equalling only 0.114 on average across observers ($SEM = 0.0334$). Such a low concentration parameter value suggests a very large orientation bandwidth. Nevertheless, AM-induced suppression is tuned significantly. As mentioned earlier, if the suppressive effect of AM would not be tuned, the S_{comp} and S_{ref} would coincide and the relative matching contrast would equal 15%. The fact that the relative matching contrast is significantly lower than 15% for all subjects does indicate that tuning is present. AM-induced suppression of a grating response is higher when the orientation of that grating matches the inducer orientation.

We do conclude that the tuning is considerably unspecific. This is illustrated in Figure 5.9. Figure 5.9A shows the summed response function S_{comp} for the comparison grating in the AM condition for a wide range of different $k_{exc,AM}$ values (green full lines). The gray full line denotes the summed response function for the Flicker condition, while the purple full line represents the S_{ref} function in the AM condition. These two latter functions are

identical to the corresponding functions displayed in Figure 5.5. The dashed green line represents the actual S_{comp} estimated by the model, i.e., obtained using the best-fitting $k_{exc,AM} = 0.114$ in the model. It can be seen that, when tuning is completely absent (light green color, $k_{exc,AM}$ approaching zero), S_{comp} and S_{ref} functions coincide in the AM condition. For intermediate values of $k_{exc,AM}$, tuning is present but not very specific. In that case, the S_{comp} function falls in between the summed response function of the Flicker condition and the S_{ref} of the AM condition. When tuning is highly specific, with the value of $k_{exc,AM}$ approaching the value estimated in Chapter 4, S_{comp} coincides with the summed response function of the Flicker condition. In other words, response suppression is completely absent for the comparison grating because the vertical inducers will not affect any neuron that is tuned to the orthogonal horizontal comparison grating orientation. The fact that the maximum of the summed response function S_{comp} is closer to S_{ref} than to the summed response of the Flicker condition in the present study does show that tuning is unspecific. This is also evident from the AM orientation tuning function recovered by our model (corresponding to the best-fitting value $k_{exc,AM} = 0.114$), which is denoted by the green dashed line in Figure 5.9B showing the AM orientation tuning functions for different $k_{exc,AM}$ values.

5.3.6 Comparison with the population code model of Chapter 4

As mentioned in section 5.2.5, the population code model used in the present study is a reduced version of the original model presented in Chapter 4. Two parameters (k_{exc} and r_0) that were incorporated as free parameters in the original model were set to physiologically-plausible values as they were insufficiently constrained by the data. Furthermore, AM-induced inhibition was omitted from the current version of the model (β was constrained to zero). We compared the AIC values for both the original model as well as the current reduced model. AIC values were not significantly different (average AIC difference = -7.8120, $p > 0.05$). This indicates that constraints on the two parameters did not affect the quality of the model fit when taking model complexity into account. Moreover, the result shows that AM-induced inhibition is indeed not required to capture the data of the present study. We therefore conclude that AM-induced inhibition does

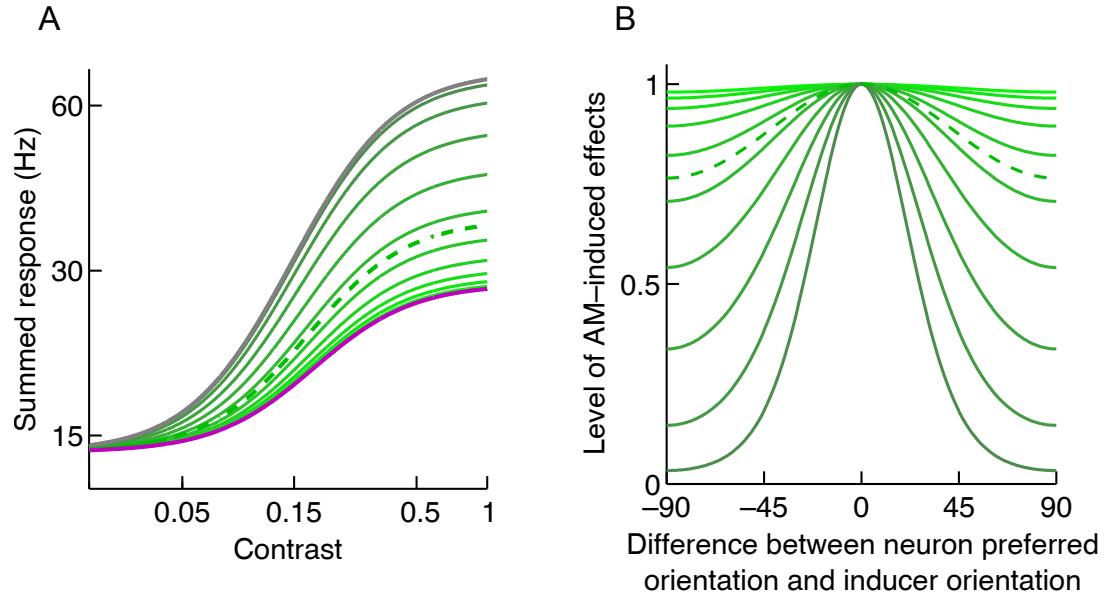


Figure 5.9. (A) Summed response functions for the comparison grating in the AM condition for a wide range of different $k_{exc,AM}$ values (green full lines). The grey full line depicts the summed response functions for the comparison and reference grating (which are identical) in the Flicker condition when $k_{exc,AM}$ is set to the estimated value of 0.114. The purple full line denotes the summed response function for the reference grating in the AM condition when $k_{exc,AM}$ equals 0.114. The green dashed line represents the summed response function for the comparison grating in the AM condition when $k_{exc,AM} = 0.114$. (B) AM orientation tuning functions for different $k_{exc,AM}$ values (green full lines). The green dashed line denotes the AM-orientation tuning function predicted by our model when $k_{exc,AM}$ equals 0.114.

not play a role in perceived grating contrast.

5.4 Discussion

In the present study, we investigated whether AM reduces the perceived contrast of a grating presented on the AM path. The starting point of our study was the finding, reported by multiple studies, that suppression of V1 responses to gratings reduces the apparent contrast of these gratings. Evidence for such a perceived contrast reduction comes from studies investigating surround masking (Snowden & Hammett, 1998; Solomon et al., 1993; Xing & Heeger, 2000; Zenger-Landolt & Heeger, 2003). When surrounded by a high-contrast grating pattern, a grating evokes smaller responses in V1 which results in a lower perceived grating contrast.

We did not apply surround masking in the current study but presented gratings in the context of AM. In Chapter 4, we found that AM significantly suppresses V1 responses to a grating presented on the AM path. As such, the AM inducers surrounding a grating stimulus may affect V1 responses in the same way as the grating masks used in surround masking experiments: both suppress or inhibit V1 responses to gratings. In both cases, suppression of responses to a grating only occurs when the low-level stimulus properties of the grating match those of the stimuli in the surrounding context. Indeed, the perceived contrast reduction observed during surround masking only occurs when the grating and surrounding pattern are similar with respect to orientation and spatial frequency (Cannon & Fullenkamp, 1991; Xing & Heeger, 2001). Likewise, AM-induced response suppression is stronger when the orientation of the grating presented on the AM path matches the inducer orientation (see Chapter 4). A match in terms of spatial frequency is probably also required for AM-induced suppression to occur, but this has not yet been investigated. In summary, as both surround masking and AM have a similar effect on V1 responses to gratings and surround masking leads to reduced perceived grating contrast, we predicted that AM also reduces perceived grating contrast.

We were able to confirm this prediction in the present study by measuring performance in a 2AFC contrast discrimination task. We used a version of the population code model developed in Chapter 4 to account for comparison choice proportion in both AM and Flicker conditions simultaneously. The model was simplified and its decoding stage was

adapted to predict contrast discrimination instead of detection performance. Notably, the model provides an excellent account of comparison choice proportion in both AM and Flicker condition. The model revealed that in the AM condition the perceived contrast of the reference grating was significantly lower than that of the comparison grating when both were presented at the reference contrast level of 15%. This was evident from the relative matching contrast derived from the model prediction, which was significantly lower than the reference grating contrast of 15%. In other words, a comparison grating had to be presented at a contrast lower than 15% in order for it to match the reference grating in terms of perceived contrast. Analysis of the contrast response functions predicted by the model revealed that AM suppresses the responses to both the reference and comparison grating. The response to the comparison grating is suppressed less strongly, however, because of the mismatch between the grating and inducer orientation. AM-induced suppression is thus orientation-tuned. As a consequence, the reference grating will evoke a lower V1 response compared to a comparison grating of the same contrast. This in turn translates to a lower perceived contrast of the reference grating relative to the comparison grating. The comparison and reference gratings can be perceived to be of equal contrast only when the comparison grating is presented at a physical contrast (i.e., the relative matching contrast) significantly lower than the reference contrast. This will lower the comparison contrast response to the level of the strongly-suppressed reference response.

An interesting aspect of our experimental design is that the reference and comparison grating are completely identical when presented at the same contrast level. Both gratings are presented on an AM path. The only difference is that the reference grating orientation matches the AM inducer orientation, while the comparison grating orientation does not. The fact that reference and comparison gratings are not perceived as equal when presented at the same contrast level can only be explained by assuming (1) that AM induces suppression and (2) that this suppression is tuned for orientation. The reference and comparison gratings presented at the same contrast level would be perceptually identical when one supposes that AM does not affect the grating response or that the difference between grating and inducer orientation is irrelevant. The relative matching

contrast would equal the reference grating contrast in that case.

In addition to suppression, we found AM to cause a small but significant level of excitation for two subjects. The effect of this excitation on discrimination performance was small compared to the suppressive effect. Multiple studies have suggested AM to induce excitation in early visual brain areas (Hidaka et al., 2011; Yantis & Nakama, 1998). Notably, this excitation is thought to interfere with stimulus representations in these brain areas. Hidaka et al., for instance, attributed the reduced detectability of gratings in the presence of AM to AM-induced excitation. We have already challenged this claim in Chapter 4, showing that the reduced grating detectability observed during AM is the result of strong AM-induced suppression. AM-induced excitation did indeed occur but it facilitated detection instead of causing interference. As mentioned in Chapter 4, this observation can be seen as an instance of the pedestal effect: the small level of excitation induced by AM acts as a pedestal, causing non-linear contrast response functions in V1 to operate in their steepest response regime (Goris et al., 2009). In this regime, small perturbations in grating contrast (on top of the pedestal) result in large and therefore easily detectable response differences.

Interestingly, we arrive at the same conclusion in the present study. AM-induced excitation, if present, has a limited effect on grating discriminability and actually improves performance. In this case, improvement is due to gratings themselves acting as a pedestal. The responses to the reference grating are situated in the steep regime of the contrast response function due to the intermediate reference contrast level. This is not the case for responses to comparison gratings of either very low or high contrast. For these extreme contrast ranges, the contrast response function is shallow. In our model, excitation is implemented through the α parameter, which causes an additive increase of the grating contrast to which V1 cells respond. More specifically, in the presence of AM-induced excitation, V1 responds to a grating of contrast c as if that grating has a contrast of $c + \alpha$. For the reference grating, the fact that contrast response functions are operating in their steepest response regime implies that even a small α value leads to a large response increase. For the comparison grating, when presented at either low or high contrast, the same α causes a much smaller response increase because of the shallowness

of the contrast response at low and high contrasts. Consequently, the pedestal effect causes an increase in perceived reference grating contrast relative to the perceived comparison grating contrast for both low and high comparison contrasts (see also Figure 1.1 in General Introduction).

It should be noted that, while excitation increases and suppression decreases responses to gratings, these effects are not contradictory and can exist simultaneously. AM-induced suppression implies a rescaling of entire V1 contrast response functions towards lower response rates. Excitation, on the other hand, has an additive rather than a scaling effect. More specifically, if no grating is presented during AM (grating contrast $c = 0$), “filling-in” occurs and V1 will respond to a virtual “filled-in” grating of contrast α . The effect of excitation will indeed be mainly visible at low contrast levels, while the scaling effect of suppression is the same at all contrast levels. As such, suppression can never fully cancel out the effect of excitation. Even though our model does not preclude excitation and suppression to exist simultaneously, the question remains why a neural mechanism would introduce both effects in the same population of neurons. We return to this issue in the General Discussion (Chapter 6).

Orientation tuning of AM-induced suppression implies that AM will introduce suppression only in neurons that are tuned to the orientation of the AM inducers. While orientation tuning was significant in the present study, the tuning was relatively unspecific compared to the specificity found in Chapter 4. During the current contrast discrimination experiment, the subpopulation of neurons affected by AM indeed contained more neurons sensitive to the inducer orientation, but also contained a lot of neurons tuned to a wide range of other orientations. This was not the case for the AM observed in Chapter 4, which involved much more specific orientation tuning. It is unclear why orientation tuning was less specific in the present study. A possible explanation may be the fact that both the horizontal and vertical inducers were presented on the same screen and alternated at the same speed. Perhaps both types of inducers were perceptually grouped horizontally into a unified single inducer. This unified inducer would consist of both orientations. As such, the induced AM may to some extent affect neurons tuned to both orientations. Orientation specificity of AM may have been reduced in this way. This hypothesis can

be readily tested by using a temporal instead of a spatial 2AFC design. In any case, we believe that a low orientation selectivity does not indicate that AM-induced effects do not occur at the level of V1. If the AM percept indeed originates and is represented at higher visual brain areas, this means that possible AM effects on responses in lower-level areas such as V1 depend on feedback connections. Neurons in higher-level areas such as MT typically have a lower orientation selectivity compared to V1 neurons (De Valois & De Valois, 1988). If higher-level areas are indeed responsible for sending feedback to V1, it is not unthinkable that these feedback signals have a low orientation selectivity, reflecting the low orientation tuning of the higher-level neurons.

We found that the estimates of the parameters specifying our model's contrast response functions were consistent with values reported in physiological monkey and cat single-cell recording studies. Especially relevant in this respect is the value of the response exponent p , which equalled 2.4 on average across subjects. The value of p reported in Chapter 4 was considerably higher, equalling 5.5 and, as such, inconsistent with physiological findings (Heeger, 1992). We attributed the high value of the response exponent to spatial uncertainty, which is known to increase the steepness of the psychometric function in contrast detection tasks (Pelli, 1985). To capture the high steepness, a high response exponent is required. In the present study, a discrimination task was used instead of a detection task. Consequently, gratings were presented at higher contrast levels (on average across trials) in the current study. This may have helped observers to better learn the exact spatial location of the grating stimuli. Arguably, it is harder to localize stimuli when they are barely visible on a relatively high number of trials. The frequent occurrence of high-contrast gratings, presented simultaneously on both sides of the screen, may have provided observers with a clear indication of the horizontal position of the gratings on a given trial. This would allow them to monitor these locations more selectively on subsequent trials. It is therefore possible that spatial uncertainty was lower in our contrast discrimination task compared to the contrast detection task of Chapter 4. This would indeed explain why the response exponent p is lower in the current study and well within the range of physiological estimates.

The formation of the AM percept presumably depends on neural processing in higher

visual areas specialized in motion, such as hMT/V5+. Neurons in these areas possess the large receptive fields required to integrate the responses to the inducer gratings, which are separated by a relatively large distance. Indeed, based on the results of Polat and Sagi (1993), we know that the V1 responses to the inducer gratings do not interact and are not integrated at the level of V1. In any case, while the AM percept itself may originate in higher-level visual areas, we do believe that the response-modulating effects (excitation, inhibition and suppression) of AM occur as early as V1. It is widely known that V1 is responsible for encoding grating contrast (Carandini, Heeger, & Movshon, 1997; Hubel & Wiesel, 1968). Psychophysical grating contrast detection and discrimination performance have been repeatedly found to mainly reflect neural responses at the level of V1 (Goris et al., 2009; Goris, Putzeys, Wagemans, & Wichmann, 2013; Putzeys et al., 2012). The fact that AM impairs performance in a contrast detection task and alters perceived contrast in a contrast discrimination task does indeed suggest that AM affects V1 responses. Perhaps more importantly, our population code model provides an accurate quantitative account of both detection and discrimination performance for a wide range of stimulus contrast levels. The low-level contrast response functions predicted by the model match those measured in V1 single-cell recording studies with respect to a wide range of parameters. This strengthens are belief that V1 is the locus of the AM effects observed in the current study.

A final remark pertains to the issue of response bias. The observed bias for choosing the comparison grating as the grating of the highest contrast should not be confused with response bias. Response bias is typically defined as a preference for certain responses irrespective of the presented stimuli. An example would be a preference to choose the grating on the left side of the screen in our discrimination task. It is important to note that such a bias could never lead to a systematic increase in comparison choice proportion because reference and comparison grating were presented randomly at the left or right side of the screen. As such, a pure response bias cannot cause a change in the relative matching contrast. The fact that observers have a preference for choosing the comparison grating is of course a bias, but not a trivial one. First of all, the bias is absent in the Flicker condition. Second, even when only considering the AM condition, there is no

reason why observers would be a-priori biased to choose the comparison grating apart from AM-induced suppression altering perceived contrast. Third, the bias seems to be contrast-dependent. As evident from Figure 5.2 for all observers, comparison choice proportions in the AM condition are higher than the corresponding choice proportions in the Flicker condition, but only for low and intermediate comparison contrast levels. At high comparison contrasts, the comparison grating is less often selected in the AM condition than in the Flicker condition. In other words, there is a cross-over in the comparison choice proportion patterns for AM and Flicker conditions (situated at a comparison contrast level of approximately 20-30%). A simple bias occurring in the AM condition, in which observers simply prefer to choose the comparison grating on a certain amount of AM condition trials, irrespective of its perceived or actual contrast, cannot account for this cross-over pattern. In the case of such a bias, comparison choice proportions are expected to be higher in the AM condition compared to the Flicker condition for all contrast levels. Our population code model, on the other hand, is able to accurately account for the cross-over pattern by only assuming AM-induced effects. Although other explanations may be possible, ours is parsimonious and therefore highly likely given the converging evidence for AM-induced effects provided in Chapter 4.

In the current study, we provided further evidence for AM-induced response suppression in V1 by investigating the effect of AM on perceived grating contrast. We started from the finding, reported by previous studies, that suppression of the V1 response to a grating lowers perceived grating contrast. As we found in Chapter 4 that AM induces strong suppression, we predicted AM to lower perceived contrast as well. This prediction was confirmed in the present study. When a grating and surrounding inducers have the same orientation, the perceived contrast of that grating is reduced relative to the perceived contrast of a grating oriented orthogonal to the surrounding inducers. The population code model developed in Chapter 4 could account for this finding and revealed that (1) AM induces suppression which lowers perceived contrast, (2) this suppression is tuned for orientation, (3) AM induces a small level of excitation for a limited number of observers and (4) this excitation facilitates discrimination.

5.5 Appendix A: Inter-subject variability

A total of 7 observers participated in the contrast discrimination task. The measured data were highly consistent for 5 observers (AW, BO, JM, ML, TV). For these observers, the population code model provided an acceptable fit to the data and model parameter estimates were relatively similar.

Most importantly, all 5 observers showed strong AM-induced suppression. The estimates of γ ranged from 75% to 80%, i.e. responses were scaled down by a factor ranging from 20% to 25%. The inter-subject variability in the γ is thus remarkably low for these observers. As mentioned earlier, the level of AM-induced excitation was less consistent across these 5 observers. Estimates of α ranged from 0% (for observer ML) to a value as high as 8% (for observer TV). Only observers JM and TV showed significant excitation, with α differing significantly from zero. It is not clear why observers differ in the amount of excitation. The results of Chapter 4 did already reveal that the level of AM-induced excitation is small and has only a minor impact on perception and on psychophysical performance. Presumably, excitation is only a minor consequence of AM and is therefore not present for all observers. On the other hand, the estimates of the parameters controlling the contrast response of the model's encoding stage were relatively consistent for the subset of 5 observers. The estimates of the response exponent p ranged from 2 to 3, while the estimates of the semi-saturation contrast c_{50} all fell between 16% and 22%. We did observe some differences in observer efficiency. The efficiency parameter was minimally 1 and maximally 2.7. This variability can be due to a wide range of observer-specific factors that are unrelated to the perception of contrast or AM and therefore not discussed here. Finally, the level of orientation tuning of the AM effects differed to some extent across the subset of 5 observers. The estimates of the concentration parameter $k_{exc,AM}$ fell between 0.05 and 0.23. Although this range seems large, all values in this range can be considered low when compared to the value of k specifying the orientation tuning functions of typical V1 excitatory receptive fields. V1 single-cell recording studies report an orientation tuning bandwidth ranging from 20° to 80°, which corresponds approximately to k values between 5.7 and 0.7, respectively. We conclude that all subjects

consistently show a low k value for the orientation tuning of AM effects, indicating broad and unspecific tuning.

There were two observers (LM and VG) that were not discussed in the main text as their results differed from the other observers. For observer LM, the population code model did provide an acceptable fit to the data ($D = 3.06, p = 0.9956$). Both data and model fit are shown in Figure 5.10A. Importantly, the level of AM-induced suppression was lower for observer LM, with γ equalling 50%. As a result of this lower suppression level, the relative matching contrast did not differ significantly from 15%. Apart from this observation, the other model parameter estimates were relatively similar to the estimates found for the other observers. When performing the contrast discrimination task, all observers were asked to focus on the central part of the screen, marked by a fixation cross. We also asked observers to pay attention to the inducer stimuli. It is possible that observer LM paid less attention to the dynamic presentation of the inducer stimuli, leading to a less strong AM percept and, consequently, less AM-induced suppression. It is also possible that this observer made more eye movements. Eye movements may disrupt the spatio-temporal structure of the AM inducer presentation, again causing a weaker AM percept. We conclude that the results of observer LM are not inconsistent with the claims made in the present study.

For the last observer, VG, the population code model did not provide an acceptable fit ($D = 31.04, p = 0.0011$). The data and model predictions are provided in Figure 5.10B. Given the fact that the model accurately predicts the performance of 6 out of 7 observers (including observer LM), we do consider observer VG to be an outlier. Perhaps this subject used a response strategy that was inconsistent across the task, across contrast levels or between AM and Flicker conditions. It can be seen that the comparison choice proportions observed in the AM and Flicker conditions did not differ considerably for observer VG. This may indicate that the AM percept was less strong or absent for observer VG, possibly for reasons that were mentioned above for observer LM. Although the fit may seem acceptable when inspecting Figure 5.10B, the predicted comparison choice proportion at a comparison contrast of 12% in the Flicker condition deviates quite substantially from the observed comparison choice proportion (see dashed rectangle in Figure 5.10B).

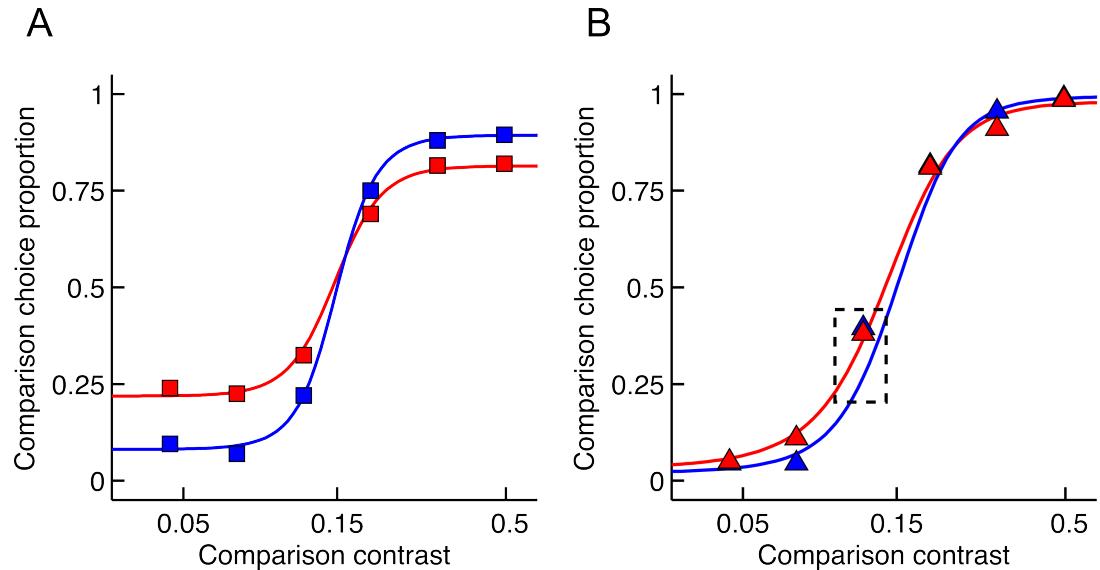


Figure 5.10. Model fit to data of observers LM (panel A) and VG (panel B). Red and blue lines and symbols represent the AM and Flicker conditions, respectively. The full lines depict the predicted comparison choice proportions of our population code model incorporating AM-induced suppression.

Furthermore, the fact that we measured 200 trials per data point implies that even small deviations from the model's predictions are indicative of a failure of the model to account for the data.

References

- Cannon, M. W., & Fullenkamp, S. C. (1991). Spatial interactions in apparent contrast: inhibitory effects among grating patterns of different spatial frequencies, spatial positions and orientations. *Vision Research*, 31(11), 1985–1998.
- Carandini, M., Heeger, D., & Movshon, J. (1997). Linearity and normalization in simple cells of the macaque primary visual cortex. *Journal of Neuroscience*, 17(21), 8621–8644.
- Chirimuuta, M., & Tolhurst, D. J. (2005). Does a Bayesian model of V1 contrast coding offer a neurophysiological account of human contrast discrimination? *Vision Research*, 45(23), 2943–2959.
- Chubb, C., Sperling, G., & Solomon, J. A. (1989). Texture interactions determine perceived contrast. *Proceedings of the National Academy of Sciences of the United States of America*, 86(23), 9631–9635.
- De Valois, R. L., & De Valois, K. K. (1988). *Spatial vision*. Oxford University Press, USA.
- Goris, R. L., Putzeys, T., Wagemans, J., & Wichmann, F. A. (2013). A neural population model for visual pattern detection. *Psychological Review*, 120(3), 472–496.
- Goris, R. L., Wichmann, F., & Henning, G. (2009). A neurophysiologically plausible population code model for human contrast discrimination. *Journal of Vision*, 9(7), 1–22.
- Haun, A. M., & Peli, E. (2013). Perceived contrast in complex images. *Journal of Vision*, 13(13), 1–21.
- Heeger, D. (1992). Half-squaring in responses of cat striate cells. *Visual Neuroscience*, 9(5), 427–443.
- Hidaka, S., Nagai, M., Sekuler, A. B., Bennett, P. J., & Gyoba, J. (2011). Inhibition of target detection in apparent motion trajectory. *Journal of Vision*, 11(10), 1–12.
- Hubel, D., & Wiesel, T. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology*, 195(1), 215–243.
- Pelli, D. G. (1985). Uncertainty explains many aspects of visual contrast detection and discrimination. *Journal of the Optical Society of America A, Optics and Image Science*, 2(9), 1508–1532.
- Polat, U., & Sagi, D. (1993). Lateral interactions between spatial channels: suppression and facilitation revealed by lateral masking experiments. *Vision Research*, 33(7), 993–999.
- Putzeys, T., Bethge, M., Wichmann, F., Wagemans, J., & Goris, R. (2012). A new perceptual bias reveals suboptimal population decoding of sensory responses. *PLoS Computational Biology*, 8(4), e1002453.
- Snowden, R. J., & Hammett, S. T. (1998). The effects of surround contrast on contrast thresholds, perceived contrast and contrast discrimination. *Vision Research*, 38(13), 1935–1945.
- Solomon, J. A., Sperling, G., & Chubb, C. (1993). The lateral inhibition of perceived contrast is indifferent to on-center/off-center segregation, but specific to orientation. *Vision Research*, 33(18), 2671–2683.

- Souto, D., & Johnston, A. (2012). Masking and color inheritance along the apparent motion path. *Journal of Vision*, 12(7), 1–18.
- Vogels, R., Spileers, W., & Orban, G. A. (1989). The response variability of striate cortical neurons in the behaving monkey. *Experimental Brain Research*, 77(2), 432–436.
- Wichmann, F. A., & Hill, N. J. (2001a). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception and Psychophysics*, 63(8), 1293–1313.
- Wichmann, F. A., & Hill, N. J. (2001b). The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Perception and Psychophysics*, 63(8), 1314–1329.
- Xing, J., & Heeger, D. J. (2000). Center-surround interactions in foveal and peripheral vision. *Vision Research*, 40(22), 3065–3072.
- Xing, J., & Heeger, D. J. (2001). Measurement and modeling of center-surround suppression and enhancement. *Vision Research*, 41(5), 571–583.
- Yantis, S., & Nakama, T. (1998). Visual interactions in the path of apparent motion. *Nature Neuroscience*, 1(6), 508–512.
- Zenger-Landolt, B., & Heeger, D. J. (2003). Response suppression in V1 agrees with psychophysics of surround masking. *Journal of Neuroscience*, 23(17), 6884–6893.

Chapter 6

General Discussion and Conclusions

6.1 Summary of results

In this doctoral dissertation, we aimed to gain more insight into the mechanisms underlying the spatio-temporal contextual modulations in low-level visual processing, either by the presence of a spatially elongated contour or the presence of an apparent motion context. In Chapter 2 and 3, we focused on the processing of local edge elements which can be grouped into a global contour. It has been proposed that perceptual as well as neurophysiological responses to such elements can be described in terms of an association field, which determines the linking strength between edge elements based on their relative position and orientation. In this model, an association strength value is typically assigned to each element in the image, indicating how likely it is that the element belongs to a contour. The association field concept has been implemented by several computational models attempting to explain how the visual system is able to extract contours embedded in a background of randomly oriented elements.

In Chapter 2, we examined how this contour integration process is supported by saccadic eye movements. To this end, we recorded observers' eye movements during a free-viewing contour search task. We found that a saliency map produced by an association field model could predict fixation locations of observers considerably above chance, indicating that observers were drawn to collinear locations in the image. As such, our results strengthened the validity of the association field model by requiring it to predict the landing positions of saccades during a contour integration task, rather than requiring it to merely predict observers' ability to detect the contour. In addition, we found that temporal and spatial properties of eye movements, reflected in fixation duration and saccade size, depended on the saliency of the contour and the reported ability to perceive a contour.

The results of the first study showed that saliency played an important role in controlling spatial and temporal aspects of saccade planning during a contour integration task. However, it remained unclear which brain processes were involved in this saccade planning, more specifically in the presaccadic selection of a target for the next fixation. In the second study reported in Chapter 3, we found that EEG activity in the interval prior to a saccade event is modulated by the presence of a contour or a region with high association

strength in peripheral vision. More specifically, we found that the presaccadic amplitude, mainly over parietal and occipital brain areas, increased with decreasing saliency of the saccade landing position. We concluded that this increased presaccadic activity may reflect increased attentional effort to compete with more salient saccade target candidates.

In Chapter 4 and Chapter 5, we examined the perceptual responses to a grating stimulus embedded in an AM context. Previous studies have found that the detectability of a stimulus presented along the AM path is impaired. This reduced detectability has generally been viewed as evidence for V1 activation along the AM path. Neural responses in V1 representing the motion path are thought to interfere with responses to physical stimuli presented along the path and as such impair the perception of these stimuli. In Chapter 4, we found that AM induced strong masking, limiting the maximum performance in detecting a target grating at high contrast levels. This masking was strongest when the target orientation was identical to the orientation of the motion-inducing stimuli. To explain these results, we developed a population code model of visual processing in V1, based on a standard contrast normalization model. We found that the predicted AM-related V1 activation was too small to either cause masking or to account for perceptual filling-in. Our model instead predicted strong suppression of low-level sensory responses along the AM path.

In Chapter 5, further evidence for AM-induced suppression of low-level responses was provided. We started from the observation made by previous studies that there exists a monotonic relationship between early visual responses and perceived contrast. If AM indeed suppresses V1 responses to a grating, the perceived contrast of a grating presented along the AM path should be reduced. To test this prediction, we measured performance in a contrast discrimination task. A population code model was developed that was similar to the model presented in Chapter 4, except that the decoder predicted discrimination performance instead of detection performance. We showed that the perceived contrast of a grating presented along an AM path was indeed lower when its orientation matched that of the motion-inducing stimuli. Our model revealed that this reduction in perceived contrast is caused by strong AM-induced suppression. A model which only included

AM-induced excitation could not account for the pattern of results.

6.2 Bottom-up saliency and eye movements

As stated in the General Introduction (Chapter 1), elements in an image are considered to be bottom-up salient when they stand out automatically from the background with respect to one or more features, without the involvement of top-down factors such as task relevance. The V1 saliency map model of Z. Li (2002) proposes that contextual modulations occurring in V1 are responsible for bottom-up saliency in an image. These contextual modulations include iso-feature suppression, which leads to decreased responses to stimuli surrounded by similar stimuli with respect to a low-level feature (Chao-Yi & Wu, 1994; DeAngelis, Freeman, & Ohzawa, 1994; Knierim & Van Essen, 1992; Levitt & Lund, 1997; Sillito, Grieve, Jones, Cudeiro, & Davis, 1995; Walker, Ohzawa, & Freeman, 1999), and collinear facilitation, which results in increased responses to collinear stimuli (Chen, Kasamatsu, Polat, & Norcia, 2001; Kapadia, Ito, Gilbert, & Westheimer, 1995; Kapadia, Westheimer, & Gilbert, 2000; W. Li, Piëch, & Gilbert, 2006; Mizobe, Polat, Pettet, & Kasamatsu, 2001; Nothdurft, Gallant, & van Essen, 1999; Polat, Mizobe, Pettet, Kasamatsu, & Norcia, 1998). According to the model, these increased responses are directly related to increased saliency. Psychophysical studies indeed suggest that observed contextual modulations in V1 provide the basis for a computation of visual salience (Jingling & Zhaoping, 2008; Koene & Zhaoping, 2007; Zhaoping & May, 2007; Zhaoping & Snowden, 2006). However, studies investigating saliency have typically considered the saliency of an image element to be reflected in subjects' reaction times in finding a particular item in a search display or detecting a change between two sequentially presented images. However, reaction times can only indirectly provide information about the spatial and temporal aspects of the search process (Zelinsky, 2008). A more direct measure of saliency can be obtained by examining the preferred image locations to which eye movements or covert attentional shifts are directed. In Chapter 2, we investigated whether collinear image structures, considered to be salient according to the association field model of Ernst et al. (2012), are likely to attract eye movements. We found that subjects visited collinear elements more than would be expected by chance, although a

considerable number of eye movements went to locations which were not salient according to the model. Furthermore, in Chapter 3, we found that presaccadic EEG responses predicted whether the upcoming saccade would land on a salient location, as predicted by an association field model.

The V1 saliency map model of Z. Li (2002) as well as the association field model of Ernst et al. (2012) assume that elements forming a collinear contour are salient due to the mutual lateral excitatory interactions between neurons responding to the collinear elements (i.e., collinear facilitation). However, a different explanation for the fact that eye movements are drawn to collinear structures is provided by the predictive coding/biased competition (PC/BC) model of Spratling (2012), by assuming an alternative mechanism for the implementation of bottom-up saliency. This model proposes that not the contour elements themselves, but the *borders* between the contour elements and the background elements are salient. The PC/BC model is an implementation of predictive coding, stating that the brain actively compares internal hypotheses of the world with bottom-up evidence to compute the prediction error. According to the PC/BC model, image locations are salient when the visual system fails to accurately predict the incoming visual information at that location. Such locations will give rise to action (e.g., making an eye movement or allocating covert attention) in order to reduce the mismatch between the reconstruction of the input and the actual input.

The activation of error-detecting neurons are assumed to reflect the degree of mismatch between the predicted and actual input. High error response values indicate that the stimulus is under-represented by the prediction and correspond to high saliency. Importantly, the model also assumes the existence of prediction neurons, which represent predictions of the underlying causes of the visual input. In a contour detection task, prediction neurons in V1 are assumed to represent the elements of the contour. The proposed interactions between these neurons are consistent with models proposing an association field of mutual excitatory interactions between neurons responding to elements lying on a contour, presumably involving long-range, intra-areal connections (Ernst et al., 2012; Z. Li, 1998). Their activity will be enhanced by top-down excitation from higher-level representations of contours activated by the presence of collinear elements in the image

(see also W. Li, Piëch, & Gilbert, 2008). Hence, the model predicts that the saliency of each contour element will be low due to low prediction error, as the visual system will be able to accurately predict the presence of a contour. However, the error in predicting elements surrounding the contour will increase so that the saliency of the border around the contour will be high. Consequently, the model predicts that eye movements and covert attention should be drawn to the border. Unfortunately, due to accuracy limits in the spatial resolution of the eye tracking system employed in Chapter 2 and Chapter 3, we were not able to differentiate between an eye movement landing either on the contour element or on the border between the contour element and the background.

Evidence providing support for the saliency of collinear contour elements comes from a change detection study of Jingling and Zhaoping (2008). In this study, observers were instructed to detect color changes in a texture stimulus defined by the orientations of element bars. In these texture stimuli, element bars are salient near texture borders due to orientation contrast. However, if collinearity contributes to bottom-up saliency, the element bars should be even more salient when they are additionally oriented parallel to the border due to collinear facilitation. The authors found indeed that changes in color were easier to detect at bars oriented parallel to the texture border, which they attributed to the additional contribution of collinear elements to saliency.

However, other findings seem to support the claim of the PC/BC model that the saliency of the elements forming a contour is low. In a visual search experiment, observers searched for a small local gap in a search display consisting of horizontal element bars, in which a vertical collinear column stood out from the background (Jingling & Tseng, 2013). The task consisted of discerning the orientation of the local target. They found that discrimination of the orientation was slower when the target was located on the contour than when it was located in the background. These results seem to indicate that the contour elements were not salient since attention did not seem to be directed to the collinear contour. However, their experimental design did not allow to examine whether it was the border surrounding the contour that was salient instead. In addition, the fact that a change in a particular location is easily detected does not necessarily indicate that this location stood out from the background; if anything, it indicates that the change

itself was salient.

The more general prediction of the PC/BC model that unpredictable stimuli are salient seems to be supported by eye movement studies showing that observers are more likely to fixate objects which are inconsistent with the general meaning of a natural scene (Loftus & Mackworth, 1978; Underwood & Foulsham, 2006; Underwood, Templeman, Lamming, & Foulsham, 2008). Moreover, it should be noted that the computation of bottom-up saliency put forward by a number of traditional saliency map models (Bruce & Tsotsos, 2009; Itti & Koch, 2001; Koch & Ullman, 1985) is not inconsistent with the saliency definition advocated by the PC/BC model. In these saliency map models, individual feature maps are constructed in parallel based on the calculation of local feature contrast, i.e., the difference between a stimulus and its surroundings with respect to a particular feature (see also General Introduction, Chapter 1). These feature maps are then integrated to produce a saliency map of the image. As mentioned above, local feature contrast is also an essential image property in determining saliency in the V1 saliency map model of Z. Li (2002). Locations with high feature contrast can be considered as locations containing high prediction error: a location is salient when the visual features at that location cannot be predicted from the surrounding features (Spratling, 2012).

6.3 Predictive coding

Predictive coding is an increasingly influential theory stating that the brain is constantly generating predictions about the external environment (Friston, 2005; Mumford, 1992; Rao & Ballard, 1999; see also General Introduction, Chapter 1). Based on these predictions, prediction errors are computed, reflecting the mismatch between the predictions and the physical sensory evidence. As stated above, the PC/BC model of Spratling (2012) provides an implementation of predictive coding theory. The PC/BC model assumes that each cortical area in the visual hierarchy contains two subpopulations of neurons: prediction neurons code the representations or predictions and error-detecting neurons code the prediction errors (see also Friston, 2005; Spratling, 2008). The computation of these residual errors can be performed by neural connections *within* a cortical area, namely by

neural interactions between error-detecting neurons and prediction neurons. Prediction neurons are enhanced through *excitatory* feedback from higher-level areas, which in turn suppress the responses of error-detecting neurons. Other implementations of predictive coding however have suggested that the calculation of residual error is performed by feedback and feedforward connections *between* cortical areas. According to these models, feedback connections from higher cortical areas carry the predictions of the visual input to lower areas, whereas feedforward pathways carry the residual errors between the predicted input and the actual lower-level input (Friston, 2005; Mumford, 1992; Rao & Ballard, 1999). The feedforward connections carrying the residual activity are assumed to be excitatory, whereas the feedback connections are assumed to be *inhibitory*, “silencing” the expected activity in lower-level areas and effectively producing residual activity.

However, studies examining the effect of predictive cortical feedback yield mixed results. Several neuroimaging and electrophysiological studies suggest that predictive feedback from higher-level cortical areas has an inhibitory effect (Alink, Schwiedrzik, Kohler, Singer, & Muckli, 2010; Den Ouden, Friston, Daw, McIntosh, & Stephan, 2009; Garrido, Kilner, Stephan, & Friston, 2009; Harrison, Stephan, Rees, & Friston, 2007; Fang, Kersten, & Murray, 2008; Murray, Kersten, Olshausen, Schrater, & Woods, 2002; Schellekens, van Wezel, Petridou, Ramsey, & Raemaekers, 2014; Todorovic, van Ede, Maris, & de Lange, 2011), while evidence from neurophysiological studies indicates that feedback connections are mostly excitatory (Hupe et al., 1998; Johnson & Burkhalter, 1997; Sandell & Schiller, 1982; Wang, Waleszczyk, Burke, & Dreher, 2000). Our findings described in Chapter 4 and Chapter 5 that V1 responses to grating stimuli are suppressed when they fit the spatio-temporal prediction of AM are consistent with an inhibitory feedback effect. We proposed that a higher-level prediction of a stimulus moving along a straight path is generated, which “explains away” lower-level responses to a physical stimulus that is consistent with this prediction.

It should be noted that excitatory feedback connections have been found to exert both excitatory and inhibitory effects (Hupe et al., 1998). The inhibitory effect on lower-level responses has been assumed to occur through activation of local inhibitory circuits (Schneider, Nelson, & Mooney, 2014; Zhang et al., 2014). This suggests the existence of

two distinct subpopulations of neurons, those which are excited and those which are indirectly inhibited through excitatory feedback. This might be consistent with predictive coding models which explicitly propose the existence of a subpopulation of prediction neurons and a subpopulation of prediction error neurons (Friston, 2005; Spratling, 2008). Feedback from higher-level areas results in a reduction of the responses of lower-level prediction error neurons but also leads to an enhanced lower-level representation of the prediction by prediction neurons. Consistent with this view, Kok, Jehee, and de Lange (2012) found that grating stimuli with an expected orientation resulted in overall reduced V1 responses, but an enhanced V1 representation of the expected orientation. These results suggest that predictive feedback suppresses the responses of prediction error neurons while “sharpening” the representation of the expected orientation.

In Chapter 4 and to a more limited extent in Chapter 5, we found evidence for the simultaneous occurrence of AM-induced excitation and inhibition. The inhibitory effect of AM is considerably stronger than the excitatory component. Indeed, in Chapter 4, the level of excitation is too small to be perceptually relevant. At a perceptual level, the main effect of AM is inhibition. Consistent with this finding, other studies provide evidence for inhibition by showing impaired perception of stimuli that can be predicted from a spatio-temporal context (He, Kersten, & Fang, 2012; Poljac, de Wit, & Wagemans, 2012; Suzuki & Cavanagh, 1995). Nevertheless, it is not clear why we also find AM-induced excitation. It may be tempting to explain this simultaneous excitation and inhibition by assuming the aforementioned predictive coding proposal of two subpopulations of neurons. However, we refrained from doing so for several reasons. First, our population code model can account for all behavioral results, thereby assuming one uniform population of V1 neurons. In our model, AM-induced excitation and inhibition (i.e., suppression) are affecting the same neurons. Second, there is currently no *direct* physiological evidence for the existence of dedicated prediction and error-detecting neurons. Third, the post-hoc assumption of such subpopulations can lead to unfalsifiable hypotheses. Future physiological studies are required to unambiguously demonstrate the existence of two functionally distinct subtypes of V1 neurons.

6.4 Conclusion

In this doctoral dissertation, the central theme was the investigation of the modulation of low-level visual processing by the spatial and temporal context. We have studied contextual influences both on a small spatial scale (collinear facilitation) as well as on a larger scale (apparent motion). Collinear facilitation forms the building block of the association field model which predicts the saliency of local features relative to their surrounding context during contour integration. We demonstrated the importance of this saliency in guiding spatial and temporal aspects of eye movements during contour integration. In addition, we found an electrophysiological correlate of saliency prior to the onset of an eye movement during a contour integration task. The second contextual effect that was investigated in this doctoral dissertation is apparent motion. We were able to refute a claim made by multiple previous studies that apparent motion alters perception by inducing excitation. Instead, computational modelling using a physiological plausible population code model revealed that apparent motion induces strong suppression. As expected, in addition to impairing stimulus detection, this suppression also alters stimulus appearance by reducing perceived contrast. Predictive coding theory can possibly provide a unifying framework in which vision is viewed as an inference process, deriving predictions from the current spatial and temporal context based on prior expectations.

References

- Alink, A., Schwiedrzik, C. M., Kohler, A., Singer, W., & Muckli, L. (2010). Stimulus predictability reduces responses in primary visual cortex. *Journal of Neuroscience*, 30(8), 2960–2966.
- Bruce, N. D., & Tsotsos, J. K. (2009). Saliency, attention, and visual search: An information theoretic approach. *Journal of vision*, 9(3), 1–24.
- Chao-Yi, L., & Wu, L. (1994). Extensive integration field beyond the classical receptive field of cat's striate cortical neurons – classification and tuning properties. *Vision Research*, 34(18), 2337–2355.
- Chen, C.-C., Kasamatsu, T., Polat, U., & Norcia, A. M. (2001). Contrast response characteristics of long-range lateral interactions in cat striate cortex. *Neuroreport*, 12(4), 655–661.
- DeAngelis, G. C., Freeman, R. D., & Ohzawa, I. (1994). Length and width tuning of neurons in the cat's primary visual cortex. *Journal of Neurophysiology*, 71(1), 347–374.
- Den Ouden, H. E., Friston, K. J., Daw, N. D., McIntosh, A. R., & Stephan, K. E. (2009). A dual role for prediction error in associative learning. *Cerebral Cortex*, 19(5), 1175–1185.
- Ernst, U. A., Mandon, S., Schinkel-Bielefeld, N., Neitzel, S. D., Kreiter, A. K., & Pawelzik, K. R. (2012). Optimality of human contour integration. *PLoS Computational Biology*, 8(5), e1002520.
- Fang, F., Kersten, D., & Murray, S. O. (2008). Perceptual grouping and inverse fMRI activity patterns in human visual cortex. *Journal of Vision*, 8(7), 2–9.
- Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1456), 815–836.
- Garrido, M. I., Kilner, J. M., Stephan, K. E., & Friston, K. J. (2009). The mismatch negativity: a review of underlying mechanisms. *Clinical Neurophysiology*, 120(3), 453–463.
- Harrison, L., Stephan, K., Rees, G., & Friston, K. (2007). Extra-classical receptive field effects measured in striate cortex with fMRI. *Neuroimage*, 34(3), 1199–1208.
- He, D., Kersten, D., & Fang, F. (2012). Opposite modulation of high-and low-level visual aftereffects by perceptual grouping. *Current Biology*, 22(11), 1040–1045.
- Hupe, J., James, A., Payne, B., Lomber, S., Girard, P., & Bullier, J. (1998). Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. *Nature*, 394(6695), 784–787.
- Itti, L., & Koch, C. (2001). Computational modeling of visual attention. *Nature Reviews Neuroscience*, 2(3), 194–203.
- Jingling, L., & Tseng, C.-H. (2013). Collinearity impairs local element visual search. *Journal of experimental psychology: human perception and performance*, 39(1), 156–167.
- Jingling, L., & Zhaoping, L. (2008). Change detection is easier at texture border bars when they are parallel to the border: Evidence for V1 mechanisms of bottom-up salience. *Perception*, 37(2), 197–206.

- Johnson, R. R., & Burkhalter, A. (1997). A polysynaptic feedback circuit in rat visual cortex. *Journal of Neuroscience*, 17(18), 7129–7140.
- Kapadia, M. K., Ito, M., Gilbert, C. D., & Westheimer, G. (1995). Improvement in visual sensitivity by changes in local context: parallel studies in human observers and in V1 of alert monkeys. *Neuron*, 15(4), 843–856.
- Kapadia, M. K., Westheimer, G., & Gilbert, C. D. (2000). Spatial distribution of contextual interactions in primary visual cortex and in visual perception. *Journal of Neurophysiology*, 84(4), 2048–2062.
- Knierim, J. J., & Van Essen, D. C. (1992). Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. *Journal of Neurophysiology*, 67(4), 961–980.
- Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiology*, 4(4), 219–227.
- Koene, A. R., & Zhaoping, L. (2007). Feature-specific interactions in salience from combined feature contrasts: Evidence for a bottom-up saliency map in V1. *Journal of Vision*, 7(7), 6–6.
- Kok, P., Jehee, J. F., & de Lange, F. P. (2012). Less is more: expectation sharpens representations in the primary visual cortex. *Neuron*, 75(2), 265–270.
- Levitt, J. B., & Lund, J. S. (1997). Contrast dependence of contextual effects in primate visual cortex. *Nature*, 387(6628), 73–76.
- Li, W., Piëch, V., & Gilbert, C. D. (2006). Contour saliency in primary visual cortex. *Neuron*, 50(6), 951–962.
- Li, W., Piëch, V., & Gilbert, C. D. (2008). Learning to link visual contours. *Neuron*, 57(3), 442–451.
- Li, Z. (1998). A neural model of contour integration in the primary visual cortex. *Neural Computation*, 10(4), 903–940.
- Li, Z. (2002). A saliency map in primary visual cortex. *Trends in Cognitive Sciences*, 6(1), 9–16.
- Loftus, G. R., & Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception and Performance*, 4(4), 565–572.
- Mizobe, K., Polat, U., Pettet, M. W., & Kasamatsu, T. (2001). Facilitation and suppression of single striate-cell activity by spatially discrete pattern stimuli presented beyond the receptive field. *Visual Neuroscience*, 18(3), 377–391.
- Mumford, D. (1992). On the computational architecture of the neocortex. *Biological Cybernetics*, 66(3), 241–251.
- Murray, S. O., Kersten, D., Olshausen, B. A., Schrater, P., & Woods, D. L. (2002). Shape perception reduces activity in human primary visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 99(23), 15164–15169.
- Nothdurft, H.-C., Gallant, J. L., & van Essen, D. C. (1999). Response modulation by texture surround in primate area V1: correlates of “popout” under anesthesia. *Visual Neuroscience*, 16(1), 15–34.
- Polat, U., Mizobe, K., Pettet, M. W., Kasamatsu, T., & Norcia, A. M. (1998). Collinear stimuli regulate visual responses depending on cell’s contrast threshold. *Nature*, 391(6667), 580–584.

- Poljac, E., de Wit, L., & Wagemans, J. (2012). Perceptual wholes can reduce the conscious accessibility of their parts. *Cognition*, 123(2), 308–312.
- Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2(1), 79–87.
- Sandell, J. H., & Schiller, P. H. (1982). Effect of cooling area 18 on striate cortex cells in the squirrel monkey. *Journal of Neurophysiology*, 48(1), 38–48.
- Schellekens, W., van Wezel, R. J., Petridou, N., Ramsey, N. F., & Raemaekers, M. (2014). Predictive coding for motion stimuli in human early visual cortex. *Brain Structure and Function*, 1–12.
- Schneider, D. M., Nelson, A., & Mooney, R. (2014). A synaptic and circuit basis for corollary discharge in the auditory cortex. *Nature*, 513(7517), 189–194.
- Sillito, A., Grieve, K., Jones, H., Cudeiro, J., & Davis, J. (1995). Visual cortical mechanisms detecting focal orientation discontinuities. *Nature*, 378(6556), 492–496.
- Spratling, M. W. (2008). Predictive coding as a model of biased competition in visual attention. *Vision Research*, 48(12), 1391–1408.
- Spratling, M. W. (2012). Predictive coding as a model of the V1 saliency map hypothesis. *Neural Networks*, 26, 7–28.
- Suzuki, S., & Cavanagh, P. (1995). Facial organization blocks access to low-level features: An object inferiority effect. *Journal of Experimental Psychology: Human Perception and Performance*, 21(4), 901–913.
- Todorovic, A., van Ede, F., Maris, E., & de Lange, F. P. (2011). Prior expectation mediates neural adaptation to repeated sounds in the auditory cortex: an MEG study. *Journal of Neuroscience*, 31(25), 9118–9123.
- Underwood, G., & Foulsham, T. (2006). Visual saliency and semantic incongruity influence eye movements when inspecting pictures. *Quarterly Journal of Experimental Psychology*, 59(11), 1931–1949.
- Underwood, G., Templeman, E., Lamming, L., & Foulsham, T. (2008). Is attention necessary for object identification? Evidence from eye movements during the inspection of real-world scenes. *Consciousness and Cognition*, 17(1), 159–170.
- Walker, G. A., Ohzawa, I., & Freeman, R. D. (1999). Asymmetric suppression outside the classical receptive field of the visual cortex. *Journal of Neuroscience*, 19(23), 10536–10553.
- Wang, C., Waleszczyk, W., Burke, W., & Dreher, B. (2000). Modulatory influence of feedback projections from area 21a on neuronal activities in striate cortex of the cat. *Cerebral Cortex*, 10(12), 1217–1232.
- Zelinsky, G. J. (2008). A theory of eye movements during target acquisition. *Psychological Review*, 115(4), 787–835.
- Zhang, S., Xu, M., Kamigaki, T., Do, J. P. H., Chang, W.-C., Jenvay, S., ... Dan, Y. (2014). Long-range and local circuits for top-down modulation of visual cortex processing. *Science*, 345(6197), 660–665.
- Zhaoping, L., & May, K. A. (2007). Psychophysical tests of the hypothesis of a bottom-up saliency map in primary visual cortex. *PLoS Computational Biology*, 3(4), e62.
- Zhaoping, L., & Snowden, R. J. (2006). A theory of a saliency map in primary visual cortex (V1) tested by psychophysics of colour–orientation interference in texture segmentation. *Visual Cognition*, 14(4-8), 911–933.