Eye-movement control during scene viewing

The roles of central and peripheral vision

To the Faculty of Human Sciences at the
University of Potsdam (Department of Psychology)
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

submitted dissertation

by

Anke Cajar

Day of submission: September 30, 2016

1st reviewer: Prof. Dr. Ralf Engbert
2nd reviewer: Prof. Dr. Felix Wichmann
3rd reviewer: Prof. Dr. Lynn Huestegge
Day of oral defense: April 25, 2017

URN urn:nbn:de:kobv:517-opus4-395536
http://nbn-resolving.de/urn:nbn:de:kobv:517-opus4-395536

Abstract

Eye movements serve as a window into ongoing visual-cognitive processes and can thus be used to investigate how people perceive real-world scenes. A key issue for understanding eye-movement control during scene viewing is the roles of central and peripheral vision, which process information differently and are therefore specialized for different tasks (object identification and peripheral target selection respectively). Yet, rather little is known about the contributions of central and peripheral processing to gaze control and how they are coordinated within a fixation during scene viewing. Additionally, the factors determining fixation durations have long been neglected, as scene perception research has mainly been focused on the factors determining fixation locations.

The present thesis aimed at increasing the knowledge on how central and peripheral vision contribute to spatial and, in particular, to temporal aspects of eye-movement control during scene viewing. In a series of five experiments, we varied processing difficulty in the central or the peripheral visual field by attenuating selective parts of the spatial-frequency spectrum within these regions. Furthermore, we developed a computational model on how foveal and peripheral processing might be coordinated for the control of fixation duration.

The thesis provides three main findings. First, the experiments indicate that increasing processing demands in central or peripheral vision do not necessarily prolong fixation durations; instead, stimulus-independent timing is adapted when processing becomes too difficult. Second, peripheral vision seems to play a prominent role in the control of fixation durations, a notion also implemented in the computational model. The model assumes that foveal and peripheral processing proceed largely in parallel and independently during fixation, but can interact to modulate fixation duration. Thus, we propose that the variation in fixation durations can in part be accounted for by the interaction between central and peripheral processing. Third, the experiments indicate that saccadic behavior largely adapts to processing demands, with a bias of avoiding spatial-frequency filtered scene regions as saccade targets. We demonstrate that the observed saccade amplitude patterns reflect corresponding modulations of visual attention.

The present work highlights the individual contributions and the interplay of central and peripheral vision for gaze control during scene viewing, particularly for the control of fixation duration. Our results entail new implications for computational models and for experimental research on scene perception.

Zusammenfassung

Blickbewegungen stellen ein Fenster in aktuelle visuell-kognitive Prozesse dar und können genutzt werden um zu untersuchen wie Menschen natürliche Szenen wahrnehmen. Eine zentrale Frage ist, welche Rollen zentrales und peripheres Sehen für die Blicksteuerung in Szenen spielen, da sie Information unterschiedlich verarbeiten und auf verschiedene Aufgaben spezialisiert sind (Objektidentifikation bzw. periphere Zielauswahl). Jedoch ist kaum bekannt, welche Beiträge zentrale und periphere Verarbeitung für die Blicksteuerung in Szenen leisten und wie sie während der Fixation koordiniert werden. Des Weiteren wurden Einflussfaktoren auf Fixationsdauern bisher vernachlässigt, da die Forschung zur Szenenwahrnehmung hauptsächlich auf Einflussfaktoren auf Fixationsorte fokussiert war.

Die vorliegende Arbeit hatte zum Ziel, das Wissen über die Beiträge des zentralen und peripheren Sehens zu räumlichen, aber vor allem zu zeitlichen Aspekten der Blicksteuerung in Szenen zu erweitern. In einer Serie von fünf Experimenten haben wir die Verarbeitungsschwierigkeit im zentralen oder peripheren visuellen Feld durch die Abschwächung selektiver Raumfrequenzanteile innerhalb dieser Regionen variiert. Des Weiteren haben wir ein computationales Modell zur Koordination von fovealer und peripherer Verarbeitung für die Kontrolle von Fixationsdauern entwickelt.

Die Arbeit liefert drei Hauptbefunde. Erstens zeigen die Experimente, dass erhöhte Verarbeitungsanforderungen im zentralen oder peripheren visuellen Feld nicht zwangsläufig zu längeren Fixationsdauern führen; stattdessen werden Fixationsdauern stimulus-unabhängig gesteuert, wenn die Verarbeitung zu schwierig wird. Zweitens scheint peripheres Sehen eine entscheidene Rolle für die Kontrolle von Fixationsdauern zu spielen, eine Idee, die auch im computationalen Modell umgesetzt wurde. Das Modell nimmt an, dass foveale und periphere Verarbeitung während der Fixation weitgehend parallel und unabhängig ablaufen, aber interagieren können um Fixationsdauern zu modulieren. Wir schlagen somit vor, dass Änderungen in Fixationsdauern zum Teil auf die Interaktion von zentraler und peripherer Verarbeitung zurückgeführt werden können. Drittens zeigen die Experimente, dass räumliches Blickverhalten sich weitgehend an Verarbeitungsanforderungen anpasst und Betrachter Szenenregionen mit Raumfrequenzfilterung als Sakkadenziele vermeiden. Wir zeigen, dass diese Sakkadenamplitudeneffekte entsprechende Modulationen der visuellen Aufmerksamkeit reflektieren.

Die vorliegende Arbeit hebt die einzelnen Beiträge und das Zusammenspiel zentralen und peripheren Sehens für die Blicksteuerung in der Szenenwahrnehmung hervor, besonders für die Kontrolle von Fixationsdauern. Unsere Ergebnisse geben neue Implikationen für computationale Modelle und experimentelle Forschung zur Szenenwahrnehmung.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Vision is human's most important sense and plays a pivotal role in perceiving and interacting with the environment. Although the acquisition of visual information seems effortless, it involves a complex dynamic between perceptual, cognitive, and motor processes. Eye movements serve as a window into this complex dynamic and provide insight into ongoing visual-cognitive and attentional processes.

Why do humans move their eyes? Visual acuity is highest at the center of gaze, the fovea, and falls off rapidly toward the periphery (Wertheim, 1894). Also, more cortical neurons are devoted to the processing of foveal than peripheral information (cortical magnification; Daniel & Whitteridge, 1961; Rovamo & Virsu, 1979). Acute vision is therefore restricted to the central visual field while the periphery is rather blurry. Given these limitations of the visual system, humans, like most animals with good vision (Land, 1999), have developed a most efficient strategy for sampling relevant information from the visual environment: they move their eyes in an alternating series of fixations and saccades. Saccades are rapid ballistic eye movements that bring new regions of interest from the blurred periphery into the high-acuity fovea for further scrutiny. During saccades, visual perception is hampered, a phenomenon called saccadic suppression (Matin, 1974). New visual information is therefore only acquired during fixations, the periods between two saccades during which the eyes are relatively stable. During fixation, viewers typically need to (i) analyze the currently fixated (i.e., foveal) stimulus, (ii) use peripheral vision to select the target of the next saccade from competing regions of interest, and (iii) program the respective saccade. In the service of these tasks, the critical decisions to be made during fixation are *when* and *where* to move the eyes.

A main goal of vision research is to understand visual-cognitive processing during everyday tasks performed in the natural environment. How do you find a friend in a crowd or find your way in an unfamiliar neighborhood? As a starting point for finding answers to these kinds of questions eye-movement control has been studied in photographs of real-world scenes, which provide static two-dimensional proxies of the real environment with reduced complexity and higher experimental control (Henderson, 2011). Nonetheless, photographs of real-world scenes are complex visual stimuli for which viewing situations are less constrained than for other viewing tasks, for example, visual search in simple task environments or reading of single sentences (Findlay & Gilchrist, 2003). Therefore, a complex dynamic of eye movements evolves during scene viewing, which is why the spatial and temporal aspects of gaze control are not as well understood as in reading or visual search.

A key issue for studying eye-movement control in real-world scenes is the relative importance of central and peripheral vision. Inherently, information is processed differently in the central and peripheral visual fields, which are therefore considered as being specialized for different tasks: analysis of details and object identification in the center versus rapid reorienting and saccade target selection in the periphery (Findlay & Gilchrist, 2003). However, the consequences of this functional segregation for eye-movement control during scene viewing have hardly been studied.

The present thesis therefore investigated both the individual contributions and the interplay of central and peripheral vision for gaze control during scene viewing. The focus was especially on the decision when to move the eyes (i.e., fixation durations), which has long been neglected in favor of the question where people fixate in scenes. A series of experiments with scene degradation in central or peripheral vision was performed and a new computational model on the control of fixation duration during scene viewing was implemented. The following sections introduce key concepts and experimental findings relevant for the subsequent chapters and provide a selective review of (i) the where and when of eye-movement control during scene viewing, (ii) models of eye-movement control, and (iii) the roles of central and peripheral vision during scene viewing. Finally, the last section of this chapter provides an overview of the studies presented in chapters 2 to 4.

1.1 The where and when of eye-movement control during scene viewing

Two essential questions of scene perception research are *where* people look in scenes and *how long* people look at certain scene patches (Henderson & Hollingworth, 1999a). While the factors determining the location of fixations in scenes have been studied extensively over the past decades both with experiments and computational modeling, the factors determining the duration of fixations have only recently been coming more into focus. However, it is well-known that both fixation locations and durations are influenced by bottom-up stimulus-based factors, top-down knowledge and goal-driven factors, and oculomotor factors.

1.1.1 Where do people look in scenes?

As Buswell (1935) and Yarbus (1967) reported in their pioneering works, viewers do not just randomly look around in scenes but fixate interesting and informative regions. A large body of research has been focused on determining which low-level image characteristics attract the gaze. This direction of research was mainly stimulated by models introducing the saliency hypothesis (Itti, Koch, & Niebur, 1998; Itti & Koch, 2000; Koch & Ullman, 1985), according to which conspicuous scene patches differing from their surround in low-level image features provide interesting information and are therefore most likely to be fixated. In these models, image features like intensity, color, or edge orientation are represented in individual topographic maps, which are combined into a unified saliency map from which the image region of highest salience is chosen as the next saccade target in a winner-take-all fashion.[1] Numerous computational models predicting fixation locations from low-level image characteristics followed Itti & Koch's model (for reviews, see Borji & Itti, 2013; Tatler, Hayhoe, Land, & Ballard, 2011). As a different method, the scene statistics approach allows to determine empirically how image characteristics differ between fixated and unfixated scene patches (Baddeley & Tatler, 2006; Mannan, Ruddock, & Wooding, 1996; Nuthmann & Einhäuser, 2015; Reinagel & Zador, 1999). Findings suggest that high-frequency edges, edge density, and clutter (amongst others) predict where people fixate.

However, a mere stimulus-driven explanation of fixation locations is too limited. Pure saliency models perform surprisingly poor in predicting fixation locations in scenes (Henderson, Brockmole, Castelhano, & Mack, 2007; Tatler et al., 2011) because they ignore the large influence of high-level cognitive factors. One important cognitive influence is the task the viewer has to accomplish during scene viewing. Both Buswell (1935) and Yarbus (1967) found that different viewing instructions prior to scene inspection (e.g., free scene exploration versus estimating people's ages in the scene; Yarbus, 1967) evoke different eye-movement patterns. Castelhano, Mack, and Henderson (2009) also showed that eye movement patterns in scenes differ qualitatively under instructions of scene memorization and visual search. Furthermore, people do not necessarily fixate salient scene regions but task relevant objects, even if they are not salient (Henderson,

---

[1] To prevent the model from constantly focussing on the most salient location, a control mechanism called inhibition of return suppresses the last attended location from the saliency map, thus allowing the next most salient location to be selected as saccade target (Itti & Koch, 2000).

Malcolm, & Schandl, 2009). During visual search, viewers likely locate scene regions with features similar to the search target (Hwang, Higgins, & Pomplun, 2009), or regions that are likely to contain the target according to scene context (Neider & Zelinsky, 2006; Spotorno, Malcolm, & Tatler, 2014). Thus, scene context and semantics also influence where people fixate. Viewers preferably fixate objects that are semantically similar to the previously fixated object or to the scene's search target (Hwang, Wang, & Pomplun, 2011). Furthermore, semantically or syntactically inconsistent objects[2] receive more fixations than consistent objects (Loftus & Mackworth, 1978; Henderson, Weeks, & Hollingworth, 1999; Võ & Henderson, 2009). Several computational models predicting fixation locations by a combination of stimulus-driven and goal-driven factors have been proposed (e.g., Navalpakkam & Itti, 2005; Torralba, Oliva, Castelhano, & Henderson, 2006; Wischnewski, Belardinelli, Schneider, & Steil, 2010; for a review, see Tatler et al., 2011).

As a third factor, fixation locations in scenes are influenced by systematic and oculomotor tendencies. For example, viewers have a strong bias for fixating the center of the scene irrespective of image features and task instructions (Tatler, 2007) and for programming saccades proceeding in the same direction as the previous saccade (saccadic momentum; Smith & Henderson, 2009). Furthermore, viewers make more horizontal than vertical or oblique saccades, and more small than long saccades (Tatler & Vincent, 2008). Tatler and Vincent (2009) showed that a computational model based solely on such oculomotor biases can outperform visual saliency models in predicting fixation locations.

1.1.2 How long do people look in scenes?

In contrast to the abundant literature on where people fixate in scenes, the question how long people fixate during scene viewing has attracted considerably less attention; so far, there is only one computational model on the control of fixation duration during scene viewing (Nuthmann, Smith, Engbert, & Henderson, 2010). This is in contrast to other fields of vision like reading research where the questions of when and where to move the eyes have both been investigated extensively (Rayner, 2009). As new visual information is only acquired during fixations, fixation duration is a crucial indicator for ongoing visual-cognitive processing (Rayner, 1998). As such a measure, fixation duration has been coming more into focus recently in scene perception research (e.g.,

---

[2] Semantically inconsistent objects violate the scene context (e.g., a cow at the beach), whereas syntactically inconsistent objects violate the scene structure (e.g., a toaster in a kitchen floating in the air; Võ & Henderson, 2009).

Henderson & Pierce, 2008; Nuthmann et al., 2010; Pannasch, Schulz, & Velichovsky, 2011).

1.1.2.1 Visual-cognitive determinants of fixation duration

Mean fixation duration during scene viewing is about 300-330 ms, with considerable variability around this mean both within and across individuals. Durations can range from less than 50 ms to more than a second (Henderson & Hollingworth, 1998; Rayner, 2009).

Fixation durations are sensitive to the quality of the visual information available during fixation (Henderson & Hollingworth, 1999a). For example, fixation durations are influenced by global scene characteristics: they increase with degradation of the whole scene in terms of decreasing luminance (Loftus, 1985), decreasing contrast (Loftus, Kaufman, Nishimoto, & Ruthruff, 1992), loss of color (Ho-Phuoc, Guyader, Landragin, & Guérin-Dugué, 2012; Nuthmann & Malcolm, 2016), or low-pass and high-pass filtering[3] (Mannan, Ruddock, & Wooding, 1995, 1997). Fixation durations also increase with global visual clutter (Henderson, Chanceaux, & Smith, 2009). Nuthmann (2016) recently showed for different viewing tasks (scene memorization, preference judgment, and object search) that fixation durations are also influenced immediately by *local* image characteristics at the current fixation position: durations increase as local edge density, clutter, and the number of meaningful image segments increase and as local luminance decreases. Furthermore, Nuthmann (2016) found that local image characteristics at the upcoming fixation position can affect the duration of the current fixation. In summary, fixation durations seem to prolong with increasing processing difficulty of the stimulus due to feature degradation or increasing information density.

In addition to stimulus-based influences, cognitive factors also play a role for fixation durations. For example, the viewer's task affects fixation durations during scene viewing, as durations are typically shorter for object search than for scene memorization (Henderson et al., 1999; Mills, Hollingworth, Van der Stigchel, & Hoffman, 2011; Nuthmann et al., 2010; Võ & Henderson, 2009; but see Castelhano et al., 2009). Furthermore, first-pass gaze durations (i.e., the sum of the durations of all fixations from first entry to first exit in an object region) are longer on semantically informative than on uninformative scene regions and longer on implausible objects not matching the scene context (De Graef, Christiaens, & d'Ydewalle, 1990; Henderson et al., 1999; Loftus & Mackworth, 1978; Underwood & Foulsham, 2006; Võ & Henderson, 2009). Implausible

---

[3] Low-pass filters blur the scene by attenuating high spatial frequencies, which carry the image's fine-scale information (e.g., edges). High-pass filters attenuate low spatial frequencies, which carry the image's coarse-scale information (e.g., general object shapes). For more details, see section 1.3.3.

objects are also fixated more often. Effects on single fixation durations, however, are inconsistent.

Finally, it should be noted that, as fixation durations merely represent the time intervals between two saccades, they are inherently constrained by the operation of the oculomotor system and saccade programming. To give one example for oculomotor factors affecting fixation duration during scene viewing, saccades continuing the trajectory of the previous saccade have been shown to have a shorter latency than saccades made in another direction; consequently, fixation durations during scene viewing have been reported to be shorter for saccades continuing the trajectory of the previous saccade and to increase with the magnitude of the change in direction (Nuthmann, 2016; Smith & Henderson, 2009, 2011; Tatler & Vincent, 2008; Wilming, Harst, Schmidt & König, 2013).

1.1.2.2 Theories on the control of fixation duration

Knowing that fixation durations during scene viewing globally adjust to visual-cognitive processing demands, a question of more recent interest is to what degree those durations reflect local, immediate visual-cognitive processing difficulty of the currently fixated stimulus. This question is also of great theoretical interest for building computational models on the control of fixation durations. There are three prominent theories on the relation between stimulus processing and the control of fixation durations, originally formulated to explain fixation durations during reading: direct control, indirect control, and mixed control (Morrison, 1984; Rayner & McConkie, 1976; Rayner & Pollatsek, 1981; see also Henderson & Smith, 2009; Trukenbrod & Engbert, 2014 for theoretical overviews).

The *direct control* theory assumes that fixation duration is affected immediately by the properties of the currently fixated stimulus and that the fixation is only terminated when the fixated stimulus is processed to a sufficient degree. Thus, direct control states that fixation durations are controlled by moment-to-moment changes of the stimulus and reflect foveal processing time. Consequently, increased processing difficulty of the currently fixated stimulus should immediately result in longer fixation durations.

The theory of *indirect control*, on the other hand, assumes that fixation durations are largely independent of the current visual input. This account is motivated by the fact that the time window during which visual-cognitive processing of the currently fixated stimulus can have immediate impact on the duration of the current fixation is rather short. First, it takes about 50 ms for visual information to reach higher cognitive areas (e.g. Foxe & Simpson, 2002). Second, minimum saccade latency (i.e., the time from stimulus onset to the initiation of an eye movement toward that stimulus) is about

150–175 ms (Rayner, Slowiaczek, Clifton, & Bertera, 1983; Rayner, 1998). Given that average fixation duration during scene viewing is around 300 ms, there is not much time left for visual-cognitive processing to influence fixation duration immediately. Indirect control accounts often propose an internal random timer, which generates fixation durations at a constant rate. This timer might adjust to task demands and viewing conditions over time (thus explaining, for example, longer fixation durations for scene memorization than visual search), but is not influenced by the currently fixated stimulus.

Finally, the *mixed control* theory combines direct and indirect control, with different ideas how this might work. For example, in the sequential attention shift model for scene viewing (Henderson, 1992), the current fixation is either terminated when the currently fixated stimulus is identified or when an identification-independent internal deadline is reached. Another concept of mixed control is the assumption of an autonomous random timer generating fixation durations that can be inhibited immediately by current processing difficulties (e.g., Engbert, Nuthmann, Richter, & Kliegl, 2005; Yang & McConkie, 2001).

1.1.2.3 Evidence from display-change experiments

Recently, several studies have investigated experimentally to which degree fixation durations during scene viewing are controlled immediately by current processing demands. To this end, the scene onset delay (SOD) paradigm has been used (Shioiri, 1993; van Diepen & Wampers, 1998). With this paradigm, the scene is replaced by a mask during the saccade prior to a prespecified critical fixation and only reappears after a certain variable delay. Assuming direct control, the duration of the critical fixation should increase monotonically with the delay period.[4] However, the results of most experiments using the SOD paradigm reveal two populations of fixation durations: one population that increases linearly with the delay and one population that is unaffected by the delay. This has been taken as evidence for a mixed control of fixation durations during scene viewing (Henderson & Pierce, 2008; Henderson & Smith, 2009; Luke, Nuthmann, & Henderson, 2013; but see Pannasch et al., 2011).

Studies using saccade-contingent display changes (i.e., changing the scene during the saccade prior to a critical fixation and changing the scene back during the saccade following the critical fixation) that did not mask the scene entirely but only reduced its quality also suggest that individual fixation durations are at least partly controlled by

---

[4] The mask onset delay (MOD) paradigm is the complement method of the SOD paradigm. During critical fixations, the scene is visible at the beginning of the fixation but replaced by a mask after a variable delay until the fixation is terminated. The MOD paradigm is used to investigate how much time viewers need to process the scene.

the currently fixated stimulus. Critical fixation durations increased linearly when scenes were increasingly degraded by decreasing luminance (Henderson, Nuthmann, & Luke, 2013; Walshe & Nuthmann, 2014), low-pass filtering (Glaholt, Rayner, & Reingold, 2013; Henderson, Olejarczyk, Luke, & Schmidt, 2014), or high-pass filtering (Glaholt et al., 2013). However, there is an ongoing debate about whether immediate processing *facilitation* can also evoke a decrease in fixation durations. While Walshe and Nuthmann (2014) reported longer fixation durations when scene luminance was increased during critical fixations, Henderson et al. (2014) found shorter fixation durations when low-pass filtered scenes were presented without filtering during critical fixations.

In summary, the empirical evidence so far suggests that fixation durations during scene viewing are at least partly controlled by the currently fixated stimulus and prolong with increasing processing demands. The mixed control account has been implemented in a computational model on the control of fixation durations during scene viewing (Nuthmann et al., 2010), which inspired the computational model proposed in the present thesis.

1.2 Models of eye-movement control

To understand the complexity of the oculomotor, perceptual, and cognitive processes underlying eye-movement control, researchers can highly benefit from theoretical and computational models. Particularly computational models are powerful tools to test hypotheses and (re)develop theories, as computational models themselves produce behavior that can be observed and measured (Fum, Del Missier, & Stocco, 2007), thus enabling not only qualitative, but also quantitative hypothesis testing. According to Lewandowsky and Farrell (2011) complex systems like the human mind cannot be understood and explained by verbal theorizing and experimental data alone, but only by complementing these techniques with quantitative modeling.

The following sections introduce two models of eye-movement control that laid the groundwork for the present computational model on the control of fixation duration (see chapter 2). The first is a general theoretical model on saccade generation and the second a computational model on the control of fixation duration during scene viewing.

1.2.1 A theoretical model of saccade generation

Findlay and Walker (1999) proposed a general theoretical model framework for saccade generation that integrates temporal and spatial aspects of eye-movement control. The model assumes two separate and parallel processing streams, the when and the where stream, that command the temporal and the spatial programming of saccades respectively. Both streams comprise five processing levels, which represent lower-level automatic processes (levels 1 to 3) and higher-level cognitive influences (levels 4 and 5) on saccade generation. At the heart of the model is the conflict resolution in the fixate and move centers on level 2, the most time-consuming part of saccade generation that determines when and where the actual movement is made. The fixate and move centers of the when and where pathway, respectively, operate via a competitive push-pull interaction. When activity in the fixate center increases, activity in the move center decreases, which favors a prolongation of the current fixation. Conversely, when activity in the move center increases and activity in the fixate center falls below a certain threshold, a trigger signal opens a gate at level 1 and a spatially coded motor command is sent to the eye muscles to initiate a saccade.

The framework is compatible with neurophysiological findings on the saccadic circuitry and can account for a variety of behavioral findings on target-elicited saccades. Although never implemented as a quantitative computational model itself, Findlay and Walker's theoretical framework influenced subsequent computational models of eye-movement control (e.g. Engbert et al., 2005; Nuthmann et al., 2010; Yang & McConkie, 2001). Most importantly, the model proposed that temporal and spatial aspects of saccade programming (i.e., the decisions of when and where to move the eyes) operate on separate pathways and can therefore partly be decoupled.

1.2.2 A computational model of fixation duration in scene viewing

Computational models on eye-movement control during scene viewing have almost exclusively been focused on fixation locations and try to predict where people look in scenes; fixation duration, as one of the most important measures of visual-cognitive processing, has completely been ignored by these models. As the only computational model, CRISP (Nuthmann et al., 2010) accounts for the control of fixation durations during scene viewing.

In CRISP an autonomous timer generates variations in fixation durations. The timer is implemented as a discrete-state continuous-time random walk that accumulates activation at a certain rate until a fixed threshold is reached. Reaching the threshold triggers a new saccade program, which consists of a labile stage and a nonlabile stage and is followed by saccade execution (see Becker & Jürgens, 1979). The time interval between two saccades is defined as the fixation duration. Although the saccade timer generates variations in fixation durations autonomously, it can be inhibited at any point in time by difficulties in visual-cognitive processing. Thus, the model assumes a mixed control of fixation duration. When processing demands increase, two mechanisms can delay saccade initiation immediately. First, the rate at which the saccade timer accumulates activation decreases by a fixed value so that the timer takes longer to reach the threshold, a mechanism termed foveal inhibition (see also Engbert et al., 2005). Second, the ongoing saccade program is cancelled with a higher probability, which automatically triggers a new saccade program. Both mechanisms prolong fixation duration effectively.

Numerical simulation studies with CRISP reproduced (i) typical mean values and distributions of fixation durations observed in scene viewing, (ii) viewing task differences (visual search vs. scene memorization) for means and distributions, and (iii) effects on fixation duration distributions typically observed with scene and mask onset delays.

Taking up Findlay and Walker's (1999) proposal of different pathways for the temporal and spatial programming of saccades, the CRISP model demonstrates that modeling fixation durations independently from fixation locations is a valid approach—although CRISP performs no analysis of scene content and completely ignores fixation locations, it can account for a variety of effects observed for fixation durations in scene viewing.

1.3 The roles of central and peripheral vision

As in any other domain of vision research, a key issue for studying real-world scene perception is the roles played by central and peripheral vision, which inherently process information differently and are therefore specialized for different tasks (analysis of visual details versus saccade target selection). However, the individual contributions as well as the interplay of central and peripheral vision for eye-movement control during scene viewing have hardly been investigated. The research outlined above on where and when people move their eyes has largely treated processing across the visual field as being homogeneous and ignored the qualitative and quantitative differences between central and peripheral vision.

In the present work gaze-contingent spatial-frequency filters were used as a tool to study the contributions of central and peripheral vision to eye-movement control. The following sections review (i) the differences in information processing between central and peripheral vision, (ii) gaze-contingent display paradigms, which are used in the present work for selectively manipulating information in central or peripheral vision, and (iii) the importance of spatial frequencies and previous research on spatial-frequency filtering in scenes.

1.3.1 Information processing in central and peripheral vision

Central vision comprises foveal vision, which extends to about 1° retinal eccentricity, and parafoveal vision, which extends from 1° to 5° eccentricity. Peripheral vision beyond 5° eccentricity is the vast majority of the visual field (see Larson & Loschky, 2009). As aforementioned, visual acuity drops rapidly with increasing distance from the fovea (Wertheim, 1894). This is due to several factors, among them the decreasing density of cone receptors (Curcio, Sloan, Kalina, & Hendrickson, 1990), increasing convergence, that is, increasing number of receptors projecting on one ganglion cell (Curcio & Allen, 1990), and cortical magnification, that is, the devotion of more cortical cells to the processing of central than peripheral information (Daniel & Whitteridge, 1961; Rovamo & Virsu, 1979). From the retina up to cortical areas, foveal information is processed by a high number of cells with small receptive fields; toward the periphery, the number of cells decreases and the sizes of the receptive fields increase. This difference between fovea and periphery increases as one moves further up the cortical hierarchy (Dumoulin & Wandell, 2008; Gattass, Gross, & Sandell, 1981; Gattass, Sousa, & Gross, 1988). Therefore, processing information about shape and color as well as object identification require (near) foveal vision, as the periphery mainly processes coarse structures and motion. Thus, central and peripheral vision can be considered as being specialized for different tasks: central vision serves object identification and analysis of details, whereas peripheral vision serves rapid spatial reorienting and the selection of new extrafoveal saccade targets (Findlay & Gilchrist, 2003).

Consequently, studies on scene viewing show that near or direct fixation is necessary to identify objects, perceive visual details and encode information to visual short-term memory (Henderson & Hollingworth, 1999b; Hollingworth, Schrock, & Henderson, 2001; Velisavljević & Elder, 2008) or long-term memory (Henderson, Williams, Castelhano, & Falk, 2003; Nelson & Loftus, 1980; Hollingworth & Henderson, 2002). Yet, peripheral vision enables object detection and categorization. Even in scenes presented in far peripheral vision at 70° or 80° eccentricity, viewers can detect animals (animal absent/present; Thorpe, Gegenfurtner, Fabre-Thorpe, & Bülthoff, 2001), categorize scenes on superordinate and basic levels (e.g., natural/urban and forests/mountains respectively; Boucart, Moroni, Thibaut, Szaffarczyk, & Greene, 2013), and even categorize local objects (face, animal, or vehicle; Boucart et al., 2016). Furthermore, peripheral vision is more useful for scene gist recognition than central vision (Larson & Loschky, 2009). Thus, object detection and categorization are (at least to a certain degree) possible with peripheral vision, but for further scrutiny and detailed analysis objects have to be in or near the fovea.

The substantial differences and functional segregation of central and peripheral vision suggest that these regions also contribute individually, and maybe differently, to eye-movement control. However, when, for example, low-level visual information is degraded in the entire scene and an increase of fixation durations is observed (Ho-Phuoc et al., 2012; Loftus, 1985; Loftus et al., 1992; Mannan et al., 1995), one cannot determine whether this effect is caused by degradation of the central stimulus, the peripheral stimulus, or both. To disentangle the roles of central and peripheral vision for eye guidance, separate manipulations of the central and the peripheral stimulus are necessary. For that, gaze-contingent moving-window and moving-mask paradigms are ideal tools.

1.3.2 Gaze-contingent moving-window and moving-mask paradigms

Gaze-contingent moving-window (McConkie & Rayner, 1975) and moving-mask (Rayner & Bertera, 1979) techniques were introduced to study the amount of information that is acquired during fixations in reading. With both techniques, a window moves in synchrony with the eye as subjects read lines of text. By varying the size of the window and by masking or altering text either inside or outside the window, the amount and kind of information that can be acquired from the stimulus is experimentally controlled. The assumption is that processing (and thus eye-movement control) is disrupted to the extent that the missing information is needed for the task at hand. With the moving-window technique, text outside the window is masked so that readers can only acquire useful information from foveal and/or parafoveal vision. By varying the size of the peripheral mask and determining how this affects eye movements and task performance, the size of the perceptual span can be studied, that is, the region from which useful visual information is acquired during a single fixation (McConkie & Rayner, 1975). Conversely, with the moving-mask technique text is masked inside the window so that information acquisition is restricted to parafoveal or peripheral vision. In reading, this investigates whether word semantics can be acquired without foveal vision (Rayner & Bertera, 1979).

Both paradigms have also been applied successfully in other areas of eye-movement research like visual search (Bertera & Rayner, 2000; Cornelissen, Bruin, & Kooijman, 2005; Geisler, Perry, & Najemnik, 2006; Greene, Pollatsek, Masserang, Lee, & Rayner, 2010; Lingnau, Schwarzbach, & Vorberg, 2010; Pomplun, Reingold, & Shen, 2001), object identification (Henderson, McClure, Pierce, & Schrock, 1997), scene perception (Foulsham, Teszka, & Kingstone, 2011; Larson & Loschky, 2009; Loschky, McConkie, Yang, & Miller, 2005; Nuthmann, 2014; Parkhurst, Culurciello, & Niebur, 2000; Reingold & Loschky, 2002; Shioiri & Ikeda, 1989; van Diepen & d'Ydewalle, 2003), or navigation in virtual environments (Fortenbaugh, Hicks, Hao, & Turano, 2007). Still, surprisingly few studies have used the moving-window and moving-mask paradigms to study scene perception (Rayner, 2009). The moving-window technique has been used more often to investigate the size of the perceptual span or functional field of view (Nuthmann, 2013; Parkhurst et al., 2000; Saida & Ikeda, 1979) or the effects of peripheral scene degradation on eye-movement behavior and task performance (e.g., Foulsham et al., 2011; Loschky et al., 2005; Reingold & Loschky, 2002; van Diepen & Wampers, 1998). A number of studies investigated how much image resolution can be reduced in the (already low-resolution) periphery until gaze control and task performance are disrupted, which is an interesting question for demanding display and image applications for saving processing resources and transmission bandwidth (for a review, see Reingold, Loschky, McConkie, & Stampe, 2003). To lower peripheral image resolution, gaze-contingent spatial-frequency filters have often been applied. As described in the next section, these filters enable interesting manipulations of information processing in central and peripheral vision.

1.3.3 Spatial frequencies

Images of real-world scenes contain information at different spatial scales. Applying Fourier analysis to 2-D images, any image can be decomposed into a set of sinusoidal components (i.e., luminance gratings) of specific spatial frequency, amplitude, phase, and orientation (De Valois & De Valois, 1990). The spatial frequency represents the number of grating cycles per degree of visual angle. High spatial frequencies carry the fine-scale information of an image about abrupt spatial changes (e.g., sharp edges and contours) or surface textures, whereas low spatial frequencies carry the coarse-scale information of an image about the general shape and structure of objects or luminance and color changes. Spatial frequencies are essential for describing visual perception, because neurons in the early visual system function as spatial filters that are selectively tuned to a particular range of spatial frequencies (Campbell & Robson, 1968; De Valois, Albrecht, & Thorell, 1982; Hubel & Wiesel, 1959).

Yet, little is known about how spatial frequencies are used for gaze control during scene viewing. To investigate this question, one can attenuate (i.e., filter) selective parts of the spatial-frequency spectrum in the scene. The reasoning is that gaze control and task performance are disrupted to the extent that the attenuated spatial frequencies are important for processing. Low-pass filters attenuate spatial frequencies above a certain cutoff value, thus attenuating fine-grained information so that scenes look blurred. High-pass filters, on the other hand, attenuate spatial frequencies below a certain cutoff value, thus attenuating coarse-grained information in the scene.

Mannan and colleagues compared eye-movement behavior in unfiltered scenes with behavior in completely high-pass and low-pass filtered scenes (Mannan et al., 1995, 1997). They found shorter saccade amplitudes and longer fixation durations for filtered scenes than for unfiltered scenes, with a stronger increase in fixation durations for low-pass than high-pass filtering. As aforementioned, studies using saccade-contingent display changes also found increased fixation durations when scenes were degraded during critical fixations by low-pass filtering (Glaholt et al., 2013; Henderson et al., 2014) or high-pass filtering (Glaholt et al., 2013). The results suggest that processing is disrupted when low or high spatial frequencies are removed from the scene.

However, when filtering spatial frequencies from the entire scene, one is unable to determine in which part of the visual field the attenuated spatial frequencies are critical for processing. This is an interesting question because the central and peripheral visual fields are differentially sensitive to parts of the spatial-frequency spectrum. The high-acuity fovea is the region that is most sensitive to medium and high spatial frequencies (Hilz & Cavonius, 1974), which are crucial for object identification and the analysis of details. The low-acuity visual periphery is mostly sensitive to low spatial frequencies (Hilz & Cavonius, 1974), which are used for rapid spatial reorienting and saccade target selection. One could assume that removing exactly these frequencies in the respective part of the visual field would disrupt scene processing the most—low-pass filtering in the central visual field would strongly impair object identification and the analysis of details, whereas high-pass filtering in the peripheral visual field would strongly impair saccade target selection. Central high-pass filters and peripheral low-pass filters, on the other hand, might be less disruptive for processing.

There are rather few studies in which spatial frequencies were filtered gaze-contingently only in central or peripheral vision during scene viewing. These studies found longer fixation durations with peripheral low-pass filtering (Loschky & McConkie, 2002; Loschky et al., 2005; Nuthmann, 2013, 2014; Parkhurst et al., 2000; van Diepen & Wampers, 1998; but see Foulsham et al., 2011), with central low-pass filtering (Nuthmann, 2014), and with peripheral high-pass filtering (van Diepen & Wampers, 1998) compared with unfiltered scene viewing. The results suggest that fixation durations increase when the scene is degraded by spatial-frequency filtering in either part of the visual field. Furthermore, the studies reported that saccade amplitudes decrease with peripheral low-pass filtering (Foulsham et al., 2011; Loschky & McConkie, 2002; Loschky et al., 2005; Nuthmann, 2013, 2014; van Diepen & Wampers, 1998) and increase with central low-pass filtering (Nuthmann, 2014), indicating that saccades preferentially target unfiltered scene regions.

In sum, the studies suggest that saccadic behavior and fixation duration largely adapt to increased processing demands induced by spatial-frequency filtering in central or peripheral vision. However, the picture is far from being complete. Most studies investigated the effects of peripheral low-pass filtering; there is little research on the effects of central filtering (Nuthmann, 2014) or high-pass filtering (van Diepen & Wampers, 1998) on gaze control during scene viewing. Furthermore, only a few studies investigated the effects of filter size or filter cutoff, and the ones who did applied low-pass filters (Loschky & McConkie, 2002; Loschky et al., 2005; Nuthmann, 2013, 2014). The present work takes a first step toward filling these gaps and to increase the knowledge on how spatial frequencies are used in central and peripheral vision during scene viewing.

1.4 The present studies

The following chapters present three studies comprising a total of five experiments and a computational model to investigate the importance of central and peripheral vision for eye-movement control in real-world scenes, particularly for the control of fixation duration. In each study, scene information was degraded in the central or the peripheral visual field by gaze-contingent spatial-frequency filtering. Processing demands were varied by applying high-pass or low-pass filters. Comparing the effects of gaze-contingent filtering on temporal and spatial aspects of eye-movement control and on task performance with unfiltered scene viewing enabled us to scrutinize the degree to which the selective attenuation of spatial frequencies disrupt central or peripheral processing. The following sections summarize the motivation and results of the studies presented in chapters 2 to 4.

1.4.1 Control of fixation duration during scene viewing by interaction of foveal and peripheral processing

In chapter 2 we present an experiment and a computational model on the role of central and peripheral vision to the control of fixation duration during scene viewing. As aforementioned, foveal vision is ideally suited for processing fine-grained information (i.e., medium and high spatial frequencies) that is crucial for the analysis of visual details, whereas peripheral vision is suited for processing coarse-grained information (i.e., low spatial frequencies) that is used for spatial reorienting. In a first experiment, we investigated the consequences of this different specialization for eye-movement control. Participants inspected real-world scenes in preparation for a memory task while high or low spatial frequencies were gaze-contingently attenuated either in the central or the peripheral visual field. Central low-pass filtering and peripheral high-pass filtering were expected to disrupt processing the most, because these conditions attenuated critical spatial frequencies that the respective regions of the visual field are most sensitive to.

The results showed that mean fixation durations increased only when potentially more useful frequencies were available for foveal analysis and peripheral target selection (i.e., with central high-pass and peripheral low-pass filtering); when those frequencies were attenuated and processing was presumably disrupted the most (i.e., with central low-pass and peripheral high-pass filtering), mean fixation durations were similar to the mean duration in the unfiltered control condition. The results challenge the assumption of the direct control theory and previous experimental findings that fixation durations prolong with increasing processing difficulty of the stimulus. To account for the effects, we implemented a computational model on the control of fixation duration during scene viewing, which assumes that fixation durations are partly controlled by the interaction of central and peripheral processing.

The model draws inspiration from CRISP (Nuthmann et al., 2010) and the model framework by Findlay and Walker (1999) and is based on four core principles. First, the model assumes that an autonomous random timer generates inter-saccadic intervals (i.e., fixation durations) at some preferred mean rate. Second, the timer can be inhibited at any point in time during fixation by ongoing visual-cognitive processing. Hence, the model assumes a mixed control mechanism for fixation duration. Third, the model accounts for spatially distributed processing by the temporal activation in two spatial compartments, which represent the unfolding of foveal and peripheral processing during the course of a fixation. Fourth, activations in the foveal and peripheral compartments proceed in parallel and independently, but can interact to influence the saccade timer—they slow down the rate of the saccade timer when activation in the foveal compartment is higher than activation in the peripheral compartment. Thus, high activation in the fovea can inhibit the saccade timer and effectively prolong fixation durations, whereas high activation in the periphery can disinhibit the saccade timer. This dynamical interaction of foveal and peripheral processing draws inspiration from the interaction between the fixate and move centers in the model framework of Findlay and Walker (1999).

The model reproduced the experimental results remarkably well, replicating the mean pattern as well as the distributions of fixation durations. Although the model shares its first two principles about the saccade timer with CRISP (Nuthmann et al., 2010), it is the first computational model of fixation durations in scene viewing that accounts for spatially distributed processing and thus for differences in central and peripheral processing. The model suggests that (i) not only central, but also peripheral vision plays a critical role in the control of fixation duration and (ii) a significant amount of variance in fixation duration is due to the interaction between central and peripheral processing.

1.4.2 Coupling of attention and saccades when viewing scenes with central and peripheral degradation

In chapter 3 we investigated the distribution of attention during central and peripheral scene degradation. Numerous studies as well as our own experimental results from chapter 2 revealed a characteristic pattern of mean saccade amplitudes with central and peripheral scene degradation: compared with viewing undegraded scenes, saccade amplitudes shorten when scenes are degraded in the peripheral visual field and lengthen when scenes are degraded in the central visual field. Thus, viewers prefer undegraded scene regions as saccade targets.

This pattern of saccade amplitudes has often been linked to visual attention (e.g., Loschky & McConkie, 2002; Nuthmann, 2013), although this hypothesis has never been tested directly. Saccades are interpreted as a behavioral manifestation of attention, because they target selected peripheral locations (at the expense of other locations) to bring into the fovea for further scrutiny (Findlay, 2009). Empirically, saccades and attention have been shown to be tightly coupled (Deubel & Schneider, 1996; Hoffman & Subramaniam, 1995; Kowler, Anderson, Dosher, & Blaser, 1995). The amplitude and direction of a saccade are therefore considered as measures of attentional selection and the spatial extent of parafoveal processing (Rayner, 1998). Following this notion, shorter saccade amplitudes typically observed with peripheral scene degradation should reflect a shrinkage of the spatial extent of attention (i.e., tunnel vision), whereas longer saccade amplitudes typically observed with central scene degradation should reflect an attentional bias toward the periphery. The latter hypothesis challenges studies reporting worse performance in peripheral target detection tasks when foveal processing demands increase (e.g., Holmes, Cohen, Haith, & Morrison, 1977; Ikeda & Takeuchi, 1975; Williams, 1989), an effect that has been attributed to a shrinkage of the functional field of view because more attentional resources are required in the fovea.

The study presented in chapter 3 tested systematically whether modulations of saccade amplitudes due to varying central and peripheral processing difficulty in scenes reflect corresponding modulations of attention. In two experiments participants viewed real-world scenes with central or peripheral low-pass or high-pass filters. While exploring the scenes in preparation for a memory task, participants simultaneously detected target stimuli presented during randomly chosen fixations in the peripheral or the central visual field.

Results confirmed the hypothesis that saccade amplitudes reflect the spatial extent of attention. Peripheral filtering shortened saccade amplitudes and decreased the detectability of peripheral but not of central targets. Central filtering, on the other hand, lengthened saccade amplitudes and decreased the detectability of central but not of peripheral targets. These results suggest that peripheral filtering causes attention to be focused on the unfiltered central region, thus producing tunnel vision with a lower perceptibility to peripheral stimuli. With central filtering, attention is withdrawn from the filtered center toward the periphery, leading to a lower perceptibility to central stimuli. Thus, we found no evidence that increased foveal load shrinks the functional field of view, as suggested by previous research. Furthermore, post-hoc analyses suggest that attention during scene viewing can be allocated in a ring-shaped fashion.

In summary, the study is the first to provide evidence for the coupling of saccade amplitudes and attention with varying central and peripheral processing difficulty during scene viewing, with a smaller attentional focus during peripheral scene degradation and an attentional bias toward the periphery during central scene degradation.

1.4.3 Spatial frequency processing in the central and peripheral visual field during scene viewing

The experimental results from chapters 2 and 3 showed that saccade target selection was increasingly modulated as scene processing became more difficult: saccade amplitudes increased with central (particularly low-pass) filtering and decreased with peripheral (particularly high-pass) filtering. Fixation durations, however, did not necessarily prolong under conditions of high processing difficulty; they only prolonged when difficulty was moderately increased and the investment of more processing time was worthwhile (i.e., with central high-pass and peripheral low-pass filtering). In a final step, we tested the validity of these results for a broader parameter space. Thus, chapter 4 presents two experiments investigating whether the previously found effects of central and peripheral spatial-frequency filtering are robust and hold when filter level or filter size vary, or if new phenomena arise.

Participants inspected real-world scenes while low-pass or high-pass filters with three different filter cutoffs (i.e., filter levels) or filter sizes were applied either to the central or the peripheral visual field. The results from both experiments corroborate our previous findings, but also provide new effects. The most important results are described in the following.

First, fixation durations did not always increase with processing difficulty due to filter type, level, and size—they only increased when difficulty was moderately increased and information could still be processed in a reasonable amount of time, but were hardly affected when processing became too difficult. This is evident because (i) fixations were longer when potentially more useful frequencies were available (i.e., with central low-pass filtering and peripheral high-pass filtering), replicating results from chapters 2 and 3, (ii) fixations prolonged with increasing filter level in the periphery with low-pass filtering, but not with high-pass filtering, and (iii) fixations prolonged with increasing peripheral filter size, but were unaffected by increasing central filter size.

Second, varying processing difficulty by type, size, or level of filtering in the peripheral visual field affected fixation durations considerably and more strongly than varying processing difficulty in the central visual field did. This corroborates both our previous experimental results and the assumption of our computational model that peripheral vision plays a critical role in the control of fixation duration.

Third, also confirming previous findings, saccade amplitudes were increasingly modulated by processing difficulty: they increased with central filtering and decreased with peripheral filtering, with the effects strengthening when filters became larger or stronger or if they attenuated potentially more useful frequencies.

Finally, the experiments revealed a trade-off between temporal and spatial aspects of eye-movement behavior: mean fixation durations were often unaffected when mean saccade amplitudes were modulated by filter manipulations, and vice versa. Thus, when processing became too difficult, processing resources were preserved by adapting either saccadic selection or saccade timing, but not both.

In summary, the experiments confirm our previously found effects on saccade amplitudes and fixation durations for a broader parameter space, corroborating the robustness of the results. Saccade amplitudes are increasingly modulated as processing difficulty increases due to spatial-frequency filtering, with opposed effects for central and peripheral filtering. Fixation durations only prolong when central or peripheral processing difficulty is moderately increased; when visual-cognitive processing becomes too difficult, stimulus-independent timing is adapted. Finally, the study is the first to vary filter sizes and filter cutoffs with both low-pass and high-pass filters.

Chapter 2

Control of fixation duration during scene viewing by interaction of foveal and peripheral processing

Jochen Laubrock*, Anke Cajar*, and Ralf Engbert
University of Potsdam, Germany

* These authors contributed equally to this work.

Running head: Control of fixation duration

published 2013 in *Journal of Vision, 13*(12):11, 1–20,
doi:10.1167/13.12.11.

Abstract

Processing in our visual system is functionally segregated, with the fovea specialized in processing fine detail (high spatial frequencies) for object identification, and the periphery in processing coarse information (low frequencies) for spatial orienting and saccade target selection. Here we investigate the consequences of this functional segregation for the control of fixation durations during scene viewing. Using gaze-contingent displays, we applied high-pass or low-pass filters to either the central or the peripheral visual field and compared eye-movement patterns with an unfiltered control condition. In contrast with predictions from functional segregation, fixation durations were unaffected when the critical information for vision was strongly attenuated (foveal low-pass and peripheral high-pass filtering); fixation durations increased, however, when useful information was left mostly intact by the filter (foveal high-pass and peripheral low-pass filtering). These patterns of results are difficult to explain under the assumption that fixation durations are controlled by foveal processing difficulty. As an alternative explanation, we developed the hypothesis that the interaction of foveal and peripheral processing controls fixation duration. To investigate the viability of this explanation, we implemented a computational model with two compartments, approximating spatial aspects of processing by foveal and peripheral activations that change according to a small set of dynamical rules. The model reproduced distributions of fixation durations from all experimental conditions by variation of few parameters that were affected by specific filtering conditions.

2.1 Introduction

Our conscious perception of a continuous, fully detailed visual environment is illusive. Due to the rapid decrease of visual acuity and spatial resolution with increasing distance from the point of gaze, high-acuity vision is limited to the foveal visual field (Jones & Higgins, 1947; Wertheim, 1894), whereas the periphery looks rather blurry and coarse grained. We therefore move our eyes about three times each second via fast and jerky movements called saccades to bring the regions of interest into high-acuity foveal vision. In this way, we actively sample our visual environment (Findlay & Gilchrist, 2003).

A large field of scene perception research has been investigating the spatial characteristics of eye movements, motivating a number of computational models to predict fixation locations in scene viewing (Hwang et al., 2009; Itti et al., 1998; Itti & Koch, 2000; Kienzle, Franz, Schölkopf, & Wichmann, 2009; Parkhurst, Law, & Niebur, 2002; Torralba et al., 2006; Tsotsos et al., 1995; Wischnewski et al., 2010; for a recent review, see Borji & Itti, 2013). The temporal aspects of eye-movement control have largely been neglected; so far only one computational model exists for fixation durations during scene viewing (Nuthmann et al., 2010). The present study therefore focuses on the temporal control of eye movements during scene perception using gaze-contingent spatial frequency filtering.

Since the foveal visual field is most sensitive to high spatial frequencies (Banks, Sekuler, & Anderson, 1991; Hilz & Cavonius, 1974; Robson & Graham, 1981), it is specialized in object identification and the analysis of fine detail. The peripheral visual field, on the other hand, is most sensitive to low spatial frequencies (Banks et al., 1991; Hilz & Cavonius, 1974; Robson & Graham, 1981), and specialized in detecting transients and coarse blobs for the rapid reorienting of overt attention and the selection of new saccade targets (Findlay & Gilchrist, 2003). How is eye-movement control affected by the selective filtering of high or low spatial frequencies in the foveal[1] or peripheral visual field? The moving-window (McConkie & Rayner, 1975) and moving-mask techniques (Rayner & Bertera, 1979) permit gaze-contingent manipulation of the peripheral and the foveal visual fields, respectively, and are therefore ideally suited to investigate this question.

Previous studies that applied spatial frequency filters to peripheral scene regions found consistent effects on saccade amplitudes, but mixed effects on fixation durations. Peripheral low-pass filters, which attenuate high spatial frequencies, decreased the mean saccade amplitude compared with an unfiltered control condition (Foulsham et al., 2011; Loschky & McConkie, 2002; Loschky et al., 2005; Shioiri & Ikeda, 1989;

---

[1] Note that, in the following, the terms “foveal visual field” and “foveal filtering” actually refer to a larger part of the visual field that includes the parafovea (up to about 5° of visual angle).

van Diepen & Wampers, 1998; Nuthmann, 2014). Most studies found longer inspection times and fixation durations with peripheral low-pass filters (Loschky & McConkie, 2002; Loschky et al., 2005; van Diepen & Wampers, 1998; Nuthmann, 2014), as well as with peripheral high-pass filters (which attenuate low-spatial frequencies; van Diepen & Wampers, 1998). Foulsham et al. (2011), however, reported that both the number of fixations and fixation durations were unaffected by peripheral low-pass filtering. Thus, it is not yet evident how peripheral spatial frequency filtering modulates fixation durations. The studies clearly point out, though, that saccades are preferably programmed to unfiltered scene regions.

Van Diepen, De Graef, and d'Ydewalle (1995) and van Diepen (2001) also confirmed this for foveal degradation in black-and-white line drawings of scenes: Reduced foveal image contrast led to longer saccade amplitudes owing to a decreased amount of short intra-object saccades in the degraded foveal region. Foveal degradation also yielded longer inspection times and fixation durations (van Diepen, 2001; van Diepen et al., 1995; Nuthmann, 2014). This effect is consistent with computational models of eye-movement control that predict longer fixation durations with increased foveal processing difficulty (Engbert et al., 2005; Nuthmann et al., 2010; Reichle, Pollatsek, Fisher, & Rayner, 1998).

Taken together, previous studies suggest that image degradation leads to increased fixation durations and the spatial selection of undegraded scene regions. The effects of high-pass filtering, i.e., of selectively attenuating information matching the demands of peripheral vision, have rarely been investigated, though, as have the effects of foveal filtering in natural scenes. Also, to our knowledge, there is no study that filtered foveal and peripheral spatial frequencies independently within a single experiment. In the present study, we therefore filtered high or low spatial frequencies in the foveal or peripheral visual field, while subjects explored natural scenes in preparation for a memory test. We compared four filter conditions with an unfiltered control condition: foveal high-pass filtering, foveal low-pass filtering, peripheral high-pass filtering, and peripheral low-pass filtering.

While the theoretical focus of our study was on the control of fixation durations, effects of filtering on saccade amplitudes are also analyzed and reported. First, we expected peripheral filtering to elicit shorter saccades and foveal filtering to elicit longer saccades compared with the control condition. The attenuation of low spatial frequencies in the periphery should impede saccade target selection more than the attenuation of high spatial frequencies; for foveal filtering, we expected effects of filter type to be reversed. Second, part of the information was missing in all conditions with filtered scenes, and fixation durations should lengthen in all experimental conditions to the extent that the missing information is used in the control of fixation duration. Given the proposed importance of foveal information for the control of fixation durations, foveal filters should particularly lengthen fixation durations. Moreover, we expected the effects to increase with increasing loss of critical spatial frequency content, i.e., with foveal low-pass filters and peripheral high-pass filters.

In the second part of the manuscript, we present a new computational model for the control of fixation durations that was inspired by the experimental results. Since our data seemed most compatible with a dynamical interaction of foveal and peripheral processing, we developed a general modeling framework that permits foveal and peripheral co-activation to modulate ongoing fixation durations. It turns out that the interactive model is in good agreement with our experimental results and that fixation duration is strongly modulated by the difference between foveal and peripheral activations.

2.2 Experiment

2.2.1 Method

2.2.1.1 Participants

Participants were 11 students from the University of Potsdam (three male, mean age: 24 years), who received course credit or €7 for participation in the experiment. All participants had normal or corrected-to-normal vision. The experiment conformed to the standards set by the Declaration of Helsinki.

2.2.1.2 Apparatus

Stimuli were presented on a 20-in. Iiyama VisionMaster Pro 514 monitor at a resolution of 1024 × 768 pixels and a refresh rate of 100 Hz. A head-chin rest ensured a viewing distance of 60 cm. While viewing was binocular, gaze position of the right eye was recorded using an EyeLink 1000 tower mount system (SR Research, Ontario, Canada) at a sampling rate of 1000 Hz. Stimulus presentation was controlled using Matlab (2009b, The Mathworks, Natick, MA) and the OpenGL-based Psychophysics Toolbox (Kleiner, Brainard, & Pelli, 2007; Pelli, 1997).

2.2.1.3 Stimuli

Stimuli were 120 color photographs of natural indoor and outdoor scenes. Of these, 36 were shot in portrait format and 84 in landscape format. Two filtered versions of each image were prepared in advance, one using a high-pass and one using a low-pass filter.

Spatial frequency filtering was realized by folding the images with a quadratic kernel with a side length of 25 pixels. Filters were recursively applied seven times in a row. The low-pass filter was a flat kernel standardized to a sum of one; the high-pass filter was a Laplacian kernel, combining a positive center with a negative surround. Filter levels were chosen heuristically so that filtering was above threshold, but the remaining information looked still usable. The signal was attenuated by more than 10 dB at spatial frequencies smaller than 0.8 cycles/° for high-pass filtering and greater than 1.4 cycles/° for low-pass filtering in a comparison of the radially averaged power spectra of filtered and unfiltered images. These numbers correspond fairly well to the maximal sensitivities of magno- and parvocellular cells in the lateral geniculate nucleus (LGN), which has been estimated at 1 and 10 cycles/°, respectively (Derrington & Lennie, 1984).

For gaze-contingent presentation, foreground and background images were merged in real-time using alpha blending. For example, in the foveal low-pass condition the blurred version of the scene was used as foreground image and the original scene as background image. The mixing ratio was given by a blending function that approximates the human contrast sensitivity function (Geisler et al., 2006, equation B7). The alpha-mask was centered at gaze location, and scaled so that only the foreground image was visible at the fixation point. The peripheral image was weighted more strongly with increasing eccentricity, so that at far eccentricities only the peripheral background image was visible. The weight of the foreground image was less than one half at eccentricities greater than 2.8°.

2.2.1.4 Design and procedure

Two filter locations were combined with two filter types, resulting in four experimental conditions: foveal low-pass filtering, foveal high-pass filtering, peripheral low-pass filtering, and peripheral high-pass filtering (Figure 2.1). A control condition presented scenes unfiltered and without a gaze-contingent window. Conditions were varied within subjects and scenes; conditions and scenes were presented in random order.

Data were collected in a single 1-hr session. The eye-tracker was calibrated at the beginning of a session and after every 24 trials. A trial started with a screen showing a central fixation trigger. The stimulus was revealed after the trigger had been fixated for at least 50 ms within 1 s from trial start; otherwise, a re-calibration was performed. Each scene was presented for 15 s. Participants were instructed to inspect the scenes carefully, and answered a three-alternative question concerning the scene content with the computer mouse after each trial. For instance, the control question for the scene illustrated in Figure 2.1 was “Which color was the passenger car?” and the three alternative answers were “blue”, “gray”, or “black”.

2.2.1.5 Data preparation

Saccades were detected in the raw time series of gaze positions using a velocity-based adaptive algorithm (Engbert & Mergenthaler, 2006; Engbert & Kliegl, 2003). A total of 71 trials were removed owing to poor recording or too much data loss. Single fixations and saccades were removed if they were neighboring eye blinks or outside of the monitor area. The first and the last event of a trial were excluded from analyses as well, since they were associated with scene onset and offset. Glissades following a saccade were assigned to the saccade; if more than one glissade followed a saccade, the glissades and their adjacent fixation and saccade were removed. In total, 4,961 fixations and 2,654 saccades were removed, leaving a total amount of 51,515 fixations and 52,756 saccades for analyses.

Figure 2.1: Effects of the four filter conditions. The yellow fixation cross illustrates the current gaze position. (Upper left) Foveal low-pass filter. (Upper right) Peripheral low-pass filter. (Lower left) Foveal high-pass filter. (Lower right) Peripheral high-pass filter.

2.2.2 Results

The questions about the scene content were answered correctly in about 90 percent of the cases (see Table 2.1). The somewhat lower fraction of correct answers for foveal high-pass filtering was probably due to control questions that referred to color information in the scene, which was difficult to extract with foveal high-pass filtering.

Data were analyzed with repeated-measures analysis of variance (ANOVA), including planned comparisons with orthogonal contrasts. We specifically tested (a) the effect of filtering, averaged across all filters and compared with the control condition, (b) the effect of filter location, (c) the effect of filter type, and (d) the interaction of filter location and filter type. We also did post-hoc tests using Bonferroni correction, with alpha set to $p = 0.0125$ (0.05/4), to examine if individual filter conditions differed significantly from the control condition. Fixation durations and saccade amplitudes were log-transformed for the analyses to achieve normal distributions.

Table 2.1: Percentage of correct answers to control questions for each condition over all participants.

Condition	Correct answer (%)
Control	90.7
Foveal low-pass filtering	88.8
Foveal high-pass filtering	86.6
Peripheral low-pass filtering	92.4
Peripheral high-pass filtering	89.4

2.2.2.1 Fixation durations

The mean fixation duration over all participants was 304 ms ($SD = 174$ ms). The distributions of fixation durations (Figure 2.2) show that peripheral low-pass filtering and especially foveal high-pass filtering led to a reduced number of short fixations, while the number of long fixations was increased. The distributions for foveal low-pass and peripheral high-pass filtering, however, did not differ markedly from the distribution for the control condition.

The ANOVA confirms the distribution patterns. Fixation duration averaged across the four experimental conditions was significantly longer than the mean fixation duration of the control condition, $F(1,10) = 12.90; MSE = 0.001; p = 0.005$. There was no main effect of filter location, $F(1,10) = 1.70; MSE = 0.002; p = 0.221$, or filter type, $F(1,10) = 3.51; MSE = 0.002; p = 0.090$, but the interaction of filter location and filter type was pronounced, $F(1,10) = 32.94; MSE = 0.001; p < 0.001$. As post-hoc tests confirm, this interaction indicates that fixation durations did not differ significantly from the control condition with peripheral high-pass filtering, $t(10) = -0.48; p = 0.641$, or foveal low-pass filtering, $t(10) = -0.18; p = 0.860$, but were increased with foveal high-pass filtering, $t(1, 10) = -5.41; p < 0.001$, and peripheral low-pass filtering, $t(1, 10) = -4.23; p = 0.002$ (see Figure 2.3).

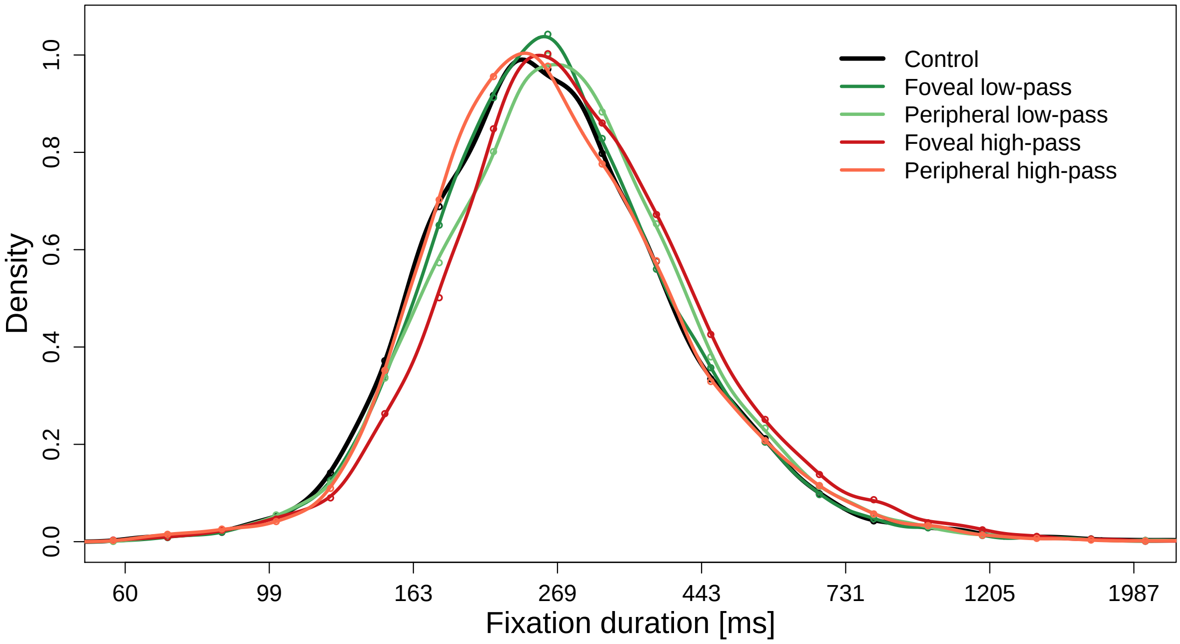

Figure 2.2: Distributions of fixation durations for the five conditions over all participants. The logarithmic scaling of the abscissa emphasizes the effects for short fixations. Lines represent kernel density estimates using a Gaussian kernel, as implemented in the R function density (bandwidth chosen according to Silverman's, 1986, equation 3.31, rule of thumb, with a weight of 1.2). The area under the curve adds up to one.

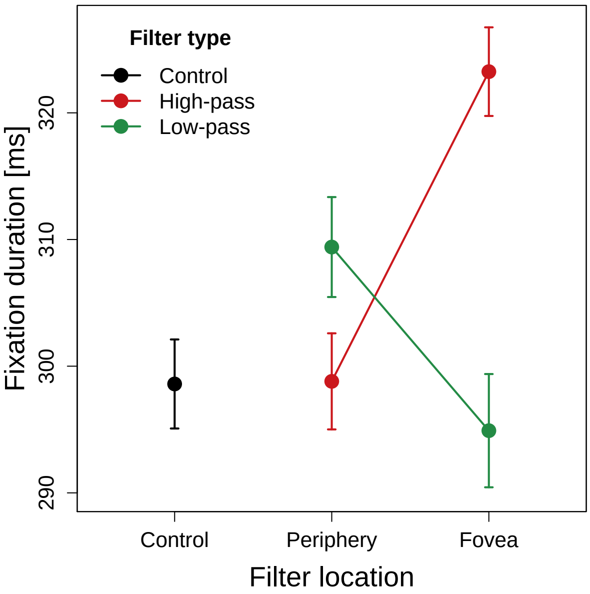

Figure 2.3: Mean fixation durations for the five conditions. Error bars represent within-subjects standard errors of the mean.

2.2.2.2 Saccade amplitudes

The mean saccade amplitude was 5.9° ($SD = 4.5°$). The distributions of saccade amplitudes (Figure 2.4) reveal that short saccades were selectively absent in both foveal filter conditions. Longer saccades occurred more frequently with foveal low-pass filtering than it did with foveal high-pass filtering or the control condition. In the peripheral filter conditions, more short and fewer long saccades occurred than in the control condition, with the pattern being more pronounced for peripheral low-pass filtering.

The ANOVA confirms these patterns, with no main effect of filter type, $F(1, 10) = 3.48; MSE = 0.004; p = 0.092$, but a pronounced effect of filter location, $F(1, 10) = 211.57; MSE = 0.005; p < 0.001$. Peripheral filtering triggered shorter and foveal filtering longer saccades than the unfiltered control condition. Although the sign of these effects was different between filter locations, their magnitude was of similar size; hence, amplitudes averaged across all filter conditions did not differ from the control condition, $F(1, 10) = 4.88; MSE = 0.006; p > 0.052$. Mean saccade amplitude differed from the control condition in three of the four filter conditions, though: peripheral low-pass filtering, $t(10) = 6.44; p < 0.001$, foveal low-pass filtering, $t(10) = -8.47; p < .001$, and foveal high-pass filtering, $t(10) = -3.25; p = 0.009$. Peripheral high-pass filtering did not significantly decrease saccade amplitude, $t(10) = 1.85; p = 0.094$. The interaction of filter type and filter location was also significant, $F(1, 10) = 49.45; MSE = 0.004; p < 0.001$—foveal filtering caused longer saccades with a low-pass filter, and peripheral filtering caused slightly longer saccades with a high-pass filter (see Figure 2.5).

2.2.3 Discussion

The effects of fixation durations, which are the focus of the present study, turned out to be strongly in conflict with hypotheses derived from the existence of a functionally segregated visual system. Compared with the control condition, fixation durations were only increased with peripheral low-pass filtering (replicating Loschky & McConkie, 2002; Loschky et al., 2005; van Diepen & Wampers, 1998) and with foveal high-pass filtering, but were unaffected with peripheral high-pass and foveal low-pass filtering. Thus, the two filter conditions with the most serious loss of information did not affect fixation durations, but the relatively more informative filter conditions did[2]. Here, the spatial frequencies that the respective region of the visual field is specialized for were still largely available if attenuated, and participants fixated longer to extract as much useful information as possible. Foveal low-pass and peripheral high-pass filtering, on the other hand, left little useful information to analyze that would make it worthwhile

---

[2] Note that although unpredicted, this result is not aberrant; we have since replicated it in other experiments with different tasks and filtering parameters (Cajar, Laubrock, & Engbert, 2013)

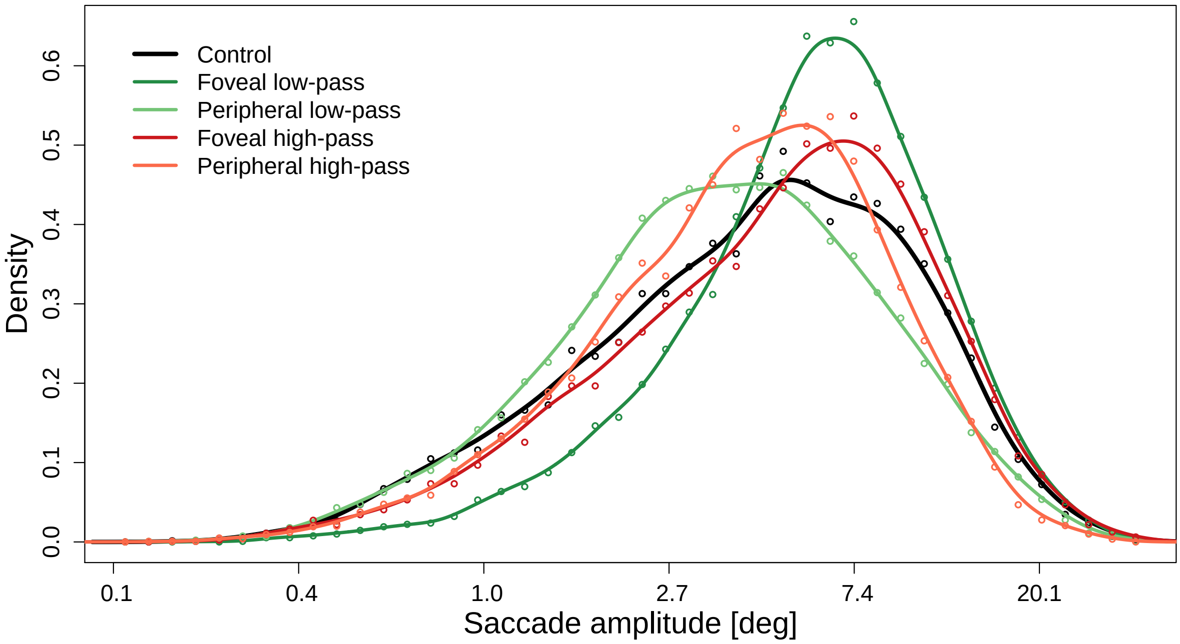

Figure 2.4: Distributions of saccade amplitudes for the five conditions over all participants. The logarithmic scaling of the abscissa emphasizes the effects for short saccades. Lines represent kernel density estimates using a Gaussian kernel, as implemented in the R function density (bandwidth chosen according to Silverman's, 1986, equation 3.31, rule of thumb, with a weight of 1.2). The area under the curve adds up to one.

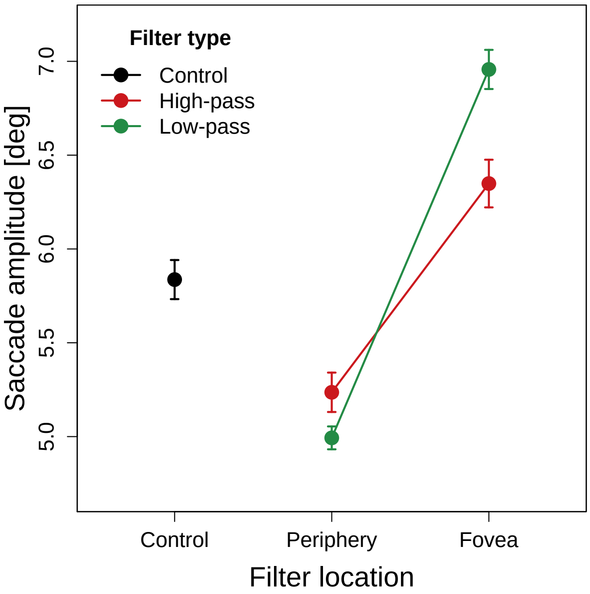

Figure 2.5: Mean saccade amplitudes for the five conditions. Error bars represent within-subjects standard errors of the mean.

to prolong fixation. Similar to the inverted-optimal viewing position (IOVP) effect in reading (Vitu, Lancelin, & Marrier d'Unienville, 2007), our results show that fixation durations increase when greater amounts of information are anticipated in a specific region.

As expected from earlier findings, the selective attenuation of spatial frequencies in the present experiment affected saccade amplitudes as well. Results were consistent with previous studies. Foveal filtering involved longer saccades with both filter types; apparently, participants preferred to explore scene regions outside the central mask, which caused a decreased number of short inspection saccades (replicating van Diepen, 2001; van Diepen et al., 1995). The amplitude effect was rather strong with foveal low-pass filtering, but less pronounced with foveal high-pass filtering. Thus, when details within the filtered region were still open for exploration with a high-pass filter, a larger proportion of short saccades remained. With peripheral filtering, on the other hand, mean saccade amplitude was shorter than it was in the control condition. This tunnel-vision effect confirms the results of previous studies (Foulsham et al., 2011; Loschky & McConkie, 2002; Shioiri & Ikeda, 1989; van Diepen & Wampers, 1998). Surprisingly, amplitudes were somewhat longer with peripheral high-pass than low-pass filtering, suggesting that high spatial frequencies were more important for the selection of peripheral saccade targets than previously assumed. This is consistent with the finding by Baddeley and Tatler (2006) that high spatial frequency content might be a reliable predictor for fixation locations.

Our findings on the influence of spatial frequency filtering on fixation durations are challenging for theoretical models of eye-movement control. One prominent concept for the control of fixation duration has been termed direct control (for a theoretical overview on direct and indirect control theories, see Henderson & Smith, 2009; Reingold, Reichle, Glaholt, & Sheridan, 2012; see also Trukenbrod & Engbert, 2014), which indicates that processing difficulty of the currently fixated stimulus immediately transfers into observed fixation durations. Under this framework, our findings conflict with the assumption that higher processing difficulty involves longer fixation durations than lower processing difficulty does, since we found (a) longer fixation durations when critical information was partially available and (b) unaffected fixation durations when critical information was strongly attenuated.

One solution to this problem might be to relate our findings to a model that implements the interaction between foveal and peripheral processing for the control of fixation duration. Such a generalization involves a very different control mechanism termed indirect control (see Trukenbrod & Engbert, 2014). While the Controlled Random-walk with Inhibition for Saccade Planning (CRISP) model (Nuthmann et al., 2010) for the control of fixation duration in scene viewing is based on indirect control, it does not address spatial aspects of processing, so that simulation studies on differences in foveal and peripheral processing are precluded. Therefore, we developed a new computational model of eye-movement control with two compartments (fovea, periphery) that interact via a small set of activation-based dynamical rules. The model is inspired by the model framework of Findlay and Walker (1999), who proposed that saccade timing and saccade target selection proceed largely independently and only interact at the lower levels of the oculomotor system.

2.3 A computational model for the control of fixation durations

For a range of visuomotor tasks, computational models of eye-movement control have been proposed within the framework of indirect control; in particular, such models successfully explained saccade timing in reading (Engbert et al., 2005), scene viewing (Nuthmann et al., 2010), and scanning tasks (Trukenbrod & Engbert, 2014). These models postulate a random saccade timer, which determines fixation durations and is modulated by ongoing visual and cognitive processing. With this approach, a variety of effects concerning mean values and distributions of fixation durations can be reproduced. The models do not distinguish between processing in the foveal and peripheral visual field, though, which is a crucial aspect in the present experiment. Therefore, we were motivated to develop a new model of eye-movement control in scene viewing with explicit assumptions on (a) foveal and peripheral processing and (b) their interaction and modulation of the random saccade timer. Such a model seems to be well constrained by our experimental data on gaze-contingent manipulations of the foveal or peripheral visual field and the resulting changes in eye-movement behavior.

2.3.1 Core principles of the model

Our model is based on four fundamental principles. First, we assume that a random timer accumulates activation toward a threshold to generate stochastic intervals between saccades, with some preferred mean value. This form of timing has been termed indirect control, since saccades are triggered autonomously, i.e., without cognitive trigger signals. There is experimental as well as theoretical support for autonomous saccade timing across a range of visuomotor tasks (e.g., Engbert & Kliegl, 2001; Engbert et al., 2005; Hooge & Erkelens, 1996, 1998; Nuthmann & Engbert, 2009; for an overview, see Trukenbrod & Engbert, 2014). Second, we assume that the autonomous random timer can be inhibited by ongoing visual-cognitive processing. Note that these two assumptions are also implemented in the CRISP model (Nuthmann et al., 2010), in which a scene-onset delay induces prolonged fixation durations by inhibition of the saccade timer. Third, we account for spatial visual processing by introducing two compartments. Both a foveal and a peripheral compartment are described by temporal activations representing the unfolding of foveal and peripheral processing over time. This choice is motivated by the functional segregation of the visual stream as outlined above. Fourth, we explicitly model the interaction of foveal and peripheral processing for the inhibition of the saccade timer. Principles three and four are qualitatively different from the CRISP model, which does not account for spatially distributed processing.

The temporal activations of the three model components, i.e., the random saccade timer, $a_T(t)$, the foveal compartment, $a_F(t)$, and the peripheral compartment, $a_P(t)$, are implemented as stochastic processes in the form of parallel, discrete-state continuous-time random walks with exponentially distributed waiting times between elementary transitions (Gillespie, 1976). Models based on random walk timing have already been very successful in explaining reaction times in simple saccadic decision tasks (see Smith & Ratcliff, 2004 for an overview).

2.3.2 A random walk for saccade timing

The random walk of the saccade timer controls the generation of the next saccade. The start of the random walk at time $t = 0$ signals the beginning of a new fixation; this state is related to an activation $a_T(t) = 0$. The random walk of activation then accumulates by increments of one toward a predefined threshold N_T with a certain rate. The time when activation reaches threshold corresponds to the fixation duration. Once the threshold is reached, a saccade program is triggered and a new fixation begins by resetting activation to a value of zero (note that for simplicity we do not model the actual saccade programming in the current version of the model; see Engbert et al., 2005, for an explicit model of saccade programming). We define the transition rate w_T for increments of the timer's random walk as

$$w_T = \frac{N_T}{t_{sac}}, \tag{2.1}$$

where N_T represents the number of states the process can adopt, and t_{sac} represents the mean duration of the timing signal.

2.3.3 Parallel processing of foveal and peripheral information

In parallel to the timer's random walk, both foveal and peripheral activations, $a_F(t)$ and $a_P(t)$ respectively, evolve over time. Here, we follow the basic framework underlying the Saccade-generation with Inhibition by Foveal Targets (SWIFT) model for reading (Engbert et al., 2005). Both foveal and peripheral activations are oculomotor variables in our model, since they directly influence saccade timing. However, they might also be interpreted psychologically in terms of visual processing, which can modulate oculomotor activation: Before processing, the foveal or peripheral stimulus is unknown; after processing, the stimulus is considered completely processed. Both cases are related to an activation of zero. During stimulus processing, activations first accumulate toward predefined thresholds, N_F or N_P for foveal and peripheral activation, respectively. During these random walk processes, activations can either increment by one unit with probability $p > 0.5$ or decrement by one unit with probability $q = 1 - p$. Each threshold can be interpreted as the maximum processing difficulty of the respective stimulus. After the threshold is reached, activation gradually declines to a value of zero, with activation either decrementing by one unit with probability p or incrementing by one unit with probability q. The transition rates for the increments of the foveal and peripheral random walks, w_F and w_P, are multiples of the timer rate w_T, with

$$w_F = j \cdot w_T \quad \text{and} \quad w_P = k \cdot w_T. \tag{2.2}$$

2.3.4 Inhibition of saccade timing by interaction of foveal and peripheral processing

Activations in the two compartments of fovea and periphery can influence the decision to start the next saccade program. This key hypothesis is motivated psychologically and neurophysiologically by Findlay and Walker's (1999) model of saccade generation by parallel processing and competitive inhibition. In this model, a hierarchy of levels of parallel processing in spatial and temporal pathways generates neural activations in fixate and move centers for the control of saccadic eye movements (Level 2 in the model by Findlay & Walker, 1999, tentatively located in the superior colliculus). Foveal and peripheral activations in our model relate to these oculomotor control centers. Therefore, the activations in the two compartments represent oculomotor tendencies to maintain fixations or to move to the next target location (i.e., the periphery). As a consequence, activations are related to cognitive processing, since higher levels of processing clearly influence neural activations in the fixate and move centers (see Findlay & Walker, 1999). The saccade timer can be considered the trigger signal for motor commands to the oculomotor muscles that generate a saccadic movement (Level 1 in the “when”-pathway of the Findlay and Walker model); it can be inhibited by upstream processing in the fixate and move centers.

Mathematically, we assume for moment-by-moment control by visual stimulus processing that the transition rate w_T of the downstream saccade timer can be modulated by the dynamical interaction between foveal and peripheral processing at any point in time during a fixation. We specifically propose that the timer rate decreases when foveal processing demands are higher than peripheral processing demands. We refer to this process as foveal inhibition (see also Engbert et al., 2005 and Nuthmann et al., 2010) and define the modulation of the timer by

$$w'_T = \frac{w_T}{h} \quad \text{with} \quad h = 1 + \rho \ \frac{\lfloor a_F(t) - a_P(t) \rfloor}{N_F}, \tag{2.3}$$

where ρ represents the strength of foveal inhibition, and $\lfloor . \rfloor$ indicates that the positive part is taken (i.e., negative values of the argument are set to zero). Thus, any kind of foveal activation leads to inhibition, as long as it is not disinhibited by a higher amount of peripheral activation at a given point in time. Peripheral activation cannot increase the transition rate of the timer, but decrease the proportion of time that inhibition is active. Numerical simulations indicated that a generalization by introducing weighting factors in the activation difference, Equation 2.3, did not improve the results for the current version of the model. The time-course of foveal and peripheral activations plays an important role for the impact on saccade timing. Foveal inhibition has the greatest effect on the saccade timer when foveal activation accumulates faster than peripheral activation during a fixation and thus leads to a positive activation difference to effectively inhibit the timer early on (large value of h). When peripheral activation rises earlier and foveal activation accumulates more slowly during a fixation, inhibition is less effective, since the timer’s random walk is a stochastic process and, consequently, can reach its threshold by chance when it is close to the threshold.

Figure 2.6 schematically illustrates the dynamics of the three model components. Since foveal and peripheral activations of our model can be interpreted as neural activations in fixate and move centers of the model by Findlay and Walker (1999), high foveal activation is in favor of a decision to prolong fixation: High foveal activation induces foveal inhibition that slows down the saccade timer and maintains the current fixation. The case of an early rise of foveal inhibition is illustrated in the upper panel of Figure 2.6. In addition to this inhibitory influence on the saccade timer from the fovea, our model proposes that high peripheral activation reduces the likelihood for foveal inhibition, effectively disinhibiting the saccade timer in favor of a move response (see Figure 2.6, lower panel). As a result, there is a dynamical interaction of foveal inhibition and peripheral disinhibition of the saccade timer; it is this interplay that determines the tendency to prolong the current fixation or to move to another fixation location by generating a saccade. Note that the interplay between foveal and peripheral activations is a specific version of Findlay and Walker's push-pull interaction between the (foveal) fixate and the (peripheral) move centers.

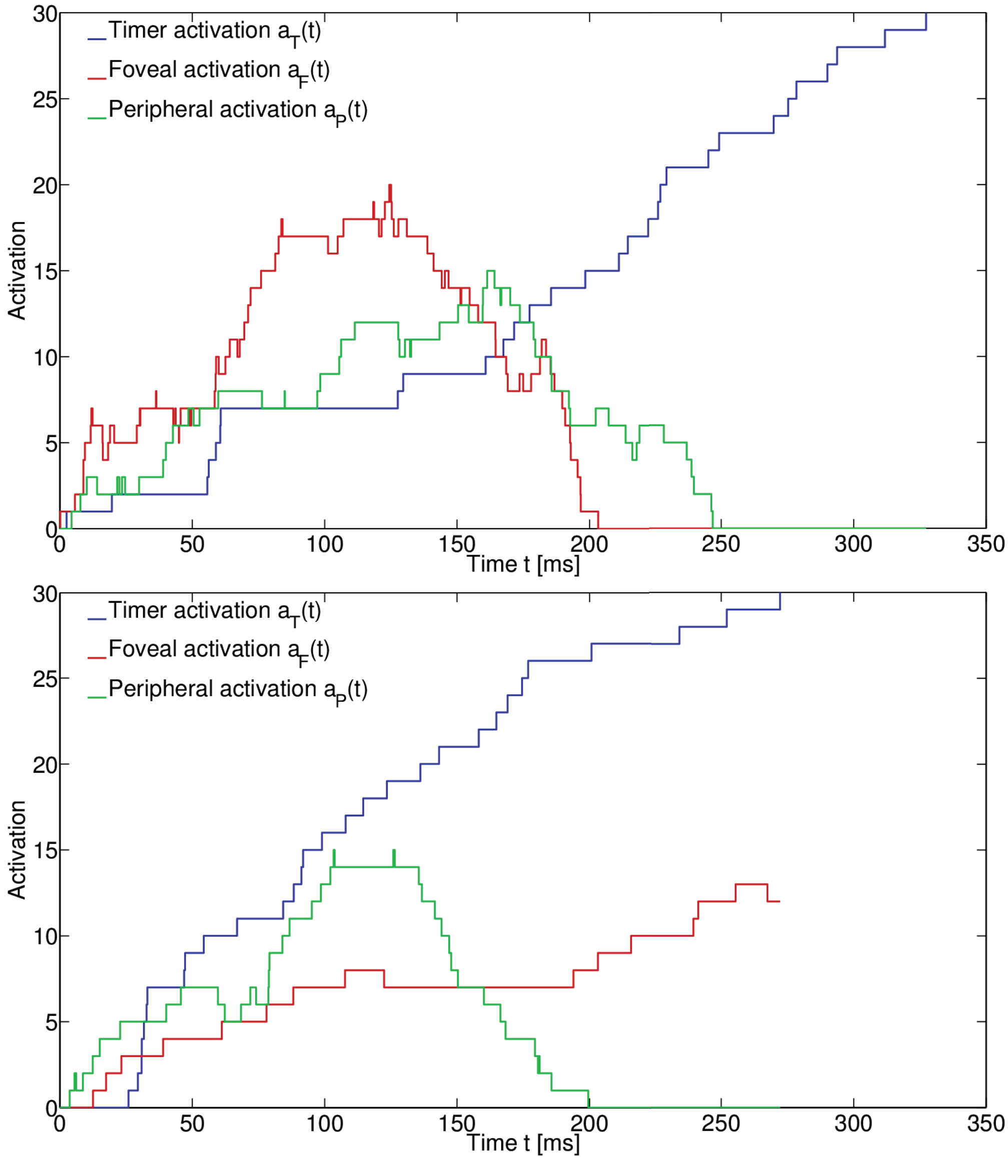

Figure 2.6: Illustration of the dynamical behavior of saccade timer (blue line), foveal compartment (red line), and peripheral compartment (green line) obtained from numerical simulation of the model. The timer accumulates activation toward a threshold that terminates the current fixation. Foveal and peripheral activations simultaneously and independently accumulate toward a maximum and then decline toward zero. The saccade timer is inhibited when foveal activation is higher than peripheral activation. (Upper panel) Foveal activation builds up early during the fixation, and foveal inhibition of the saccade timer is effective during the initial part of the fixation (duration 327 ms). (Lower panel) Peripheral activation builds up early, leading to disinhibition during the initial part of the fixation (duration 272 ms).

2.3.5 Numerical simulation procedures

To illustrate how the random walks for the timer, the foveal, and peripheral activation work in parallel, consider a system in state $S(t) = (a_T, a_F, a_P)$ at time t, which changes to an adjoined state $S(t+\tau)$ at time $t+\tau$. With each time step, only one of the three random walks changes its state, while the other two random walks remain unchanged. Table 2.2 summarizes all possible state-transitions in the model. The three transition rates each represent the probability for a specific state transition, so that the random walk with the highest rate has the highest probability of changing its state (for further details, see Gillespie, 1976). The total transition rate W is defined as the sum of the three individual rates,

$$W = w'_T + w_F + w_P. \tag{2.4}$$

The algorithm consists of two steps. First, a time step is chosen. For each random walk, the transition probability from the current state to the next state depends on the past only through the current state; this is characteristic for Markov processes (e.g., Gardiner, 2004; van Kampen, 1981). Consequently, the waiting time τ between different transitions follows an exponential distribution (Gillespie, 1976), and thus can be transformed from a uniformly distributed random number by

$$\tau = -\frac{1}{W}\, log(1-r), \tag{2.5}$$

where the inverse of the total transition rate W represents the mean waiting time in a given state $S(t)$, and r is a random number with equal probability in $0 \le r < 1$. Second, a transition is selected in proportion to the transition rate of the walks. The probabilities for selecting a transition in the saccade timer, foveal, and peripheral compartments are given by w'_T/W, w_F/W, and w_P/W, respectively.

Table 2.2: Transition events and transition rate.

Random walk	Transition from $S(t) = (a_T, a_F, a_P)$ to $S(t+\tau) = ...$			Transition rate
Saccade timer	$a_T + 1$	a_F	a_P	w'_T
Foveal activation	a_T	$a_F \pm 1^*$	a_P	w_F
Peripheral activation	a_T	a_F	$a_P \pm 1^*$	w_P

Notes: *, Foveal and peripheral activations are incremented by $+1$ before the respective threshold N_F or N_P is reached and are decremented by –1 thereafter with probability p. Activations decrease by –1 after the threshold is reached and increase by $+1$ thereafter with probability $q = 1 - p$.

2.3.6 Numerical simulation study

A simulation study was conducted to investigate if the model was able to (a) reproduce the mean values and distributions of fixation durations from the present experiment, and (b) provide a viable mechanism for the experimental findings on fixation duration effects of foveal and peripheral filtering. We thus simulated fixation durations from 11 subjects for foveal low-pass filtering, foveal high-pass filtering, peripheral low-pass filtering, peripheral high-pass filtering, and for the unfiltered control condition.

2.3.6.1 Hypotheses

We assume that selective spatial frequency filtering in the foveal or peripheral visual field modulates saccade timing by affecting the random walks for foveal or peripheral processing. Specifically, foveal filtering should affect foveal processing, and peripheral filtering should affect peripheral processing. The time course of foveal and peripheral processing can therefore be modulated in three ways. First, filtering could change the transition rates w_F and w_P for foveal and peripheral activations by changing the factors j and k (see Equation 2.2). Second, filtering could change the probability p of the random walks to increment (or decrement, after the threshold has been reached) by one unit. Third, filtering could change the values N_F and N_P for the maximum difficulty of the foveal or peripheral stimulus. We assumed that these modulations could co-occur, of course, but kept a number of model parameters fixed across conditions for psychological plausibility of the model. Since there is neurophysiological evidence for fixed thresholds but variable growth rates obtained from experiments on the generation of voluntary eye movements (Hanes & Schall, 1996), we assumed that the thresholds for foveal and peripheral activations in our model, N_F and N_P, are constant across all five filter conditions. The growth of activation, on the other hand, was supposed to change with the different filter conditions—either by changing the transition rates, or by changing the probability of the random walk to increment/decrement by one unit.

2.3.6.2 Modeling results

Numerical simulations were run to find the best-fitting model on fixation durations. We aimed at a minimal model with as few free parameters as possible, which captures all the main qualitative effects of fixation durations observed in the experimental data. A genetic algorithm (Goldberg, 1989; Holland, 1975) was used for parameter estimation. Predefined parameter ranges (see Table 2.3) were chosen for mathematical reasons or for neurophysiological or psychological plausibility. A quantile maximum likelihood approach (Heathcote, Brown, & Mewhort, 2002) served as a goodness-of-fit measure

Table 2.3: Model: Average parameter values for the best model.

Parameter	Symbol	*M*	*SE*	Predefined range
Default random walk transition rate for the saccade timer	$w_T = N_T/t_{sac}$	$N_T = 9$	0.28	3–50
		$t_{sac} = 253$ ms	5.39 ms	150–300 ms
Random walk transition rate for foveal processing	$w_F = j \cdot w_T$	$j_C = 2.61$	0.15	1–5
		$j_{fLP} = 2.58$	0.14	1–5
		$j_{fHP} = 3.44$	0.12	1–5
Random walk transition rate for peripheral processing	$w_P = k \cdot w_T$	$k_C = 3.15$	0.23	0.01–5
		$k_{pLP} = 2.53$	0.17	0.01–5
		$k_{pHP} = 3.06$	0.13	0.01–5
Maximum foveal processing difficulty	N_F	33	0.39	3–50
Maximum peripheral processing difficulty	N_P	22	1.49	3–50
Probability of foveal and peripheral random walk to increment by one	p	0.66	0.01	0.55–1.0
Strength of foveal inhibition	ρ	4.37	0.48	0.1–10

and quantified how much the simulated fixation duration distributions deviated from the experimental distributions (see Appendix for details on the parameter estimation and fitting procedure).

Explorative numerical simulations of two model variants—Model A with variable transition rates for the five filter conditions, and Model B assuming variable probabilities of incrementing/decrementing—indicated that Model A provided a qualitatively better fit to the experimental distributions. This model has 12 free parameters. Six parameters were not allowed to change across the five filter conditions—the three thresholds of the random walks, N_T, N_F, and N_P, the mean duration of the timing signal, t_{sac}, the probability p for the foveal and peripheral random walk to increment (or decrement) by one unit, and the strength of foveal inhibition, ρ. The transition rates w_F and w_P for the foveal and peripheral random walk, however, were allowed to change for the different filter conditions by changing the factor j with foveal filtering, and the factor k with peripheral filtering. Six parameters are necessary to describe this modulation, since for foveal as well as for peripheral filtering, there is one transition rate for the control condition, one for high-pass filtering, and one for low-pass filtering.

We fitted this model to the experimental data of each subject separately and obtained 11 sets of model parameters. Each parameter set was used to simulate as many observations (fixation durations) as the respective subject contributed to the experimental data set. Figure 2.7 illustrates for 1 of the 11 subjects how the simulated fixation duration distributions fit the respective experimental distributions. The model captured the shape for each experimental distribution well, with the characteristic positive skew including a longer tail and the mode below the mean. Despite individual differences between the subjects, the model captured the distributions of all subjects equally well. Although only the distributions were fitted, Figure 2.8 illustrates that the simulated data set combined from all 11 subjects recovered the experimental pattern of mean fixation durations remarkably well; only the mean value for foveal high-pass filtering was slightly underestimated by the model. Additionally, the simulation of fixation durations with the average values of the model parameters across subjects yielded very similar results.

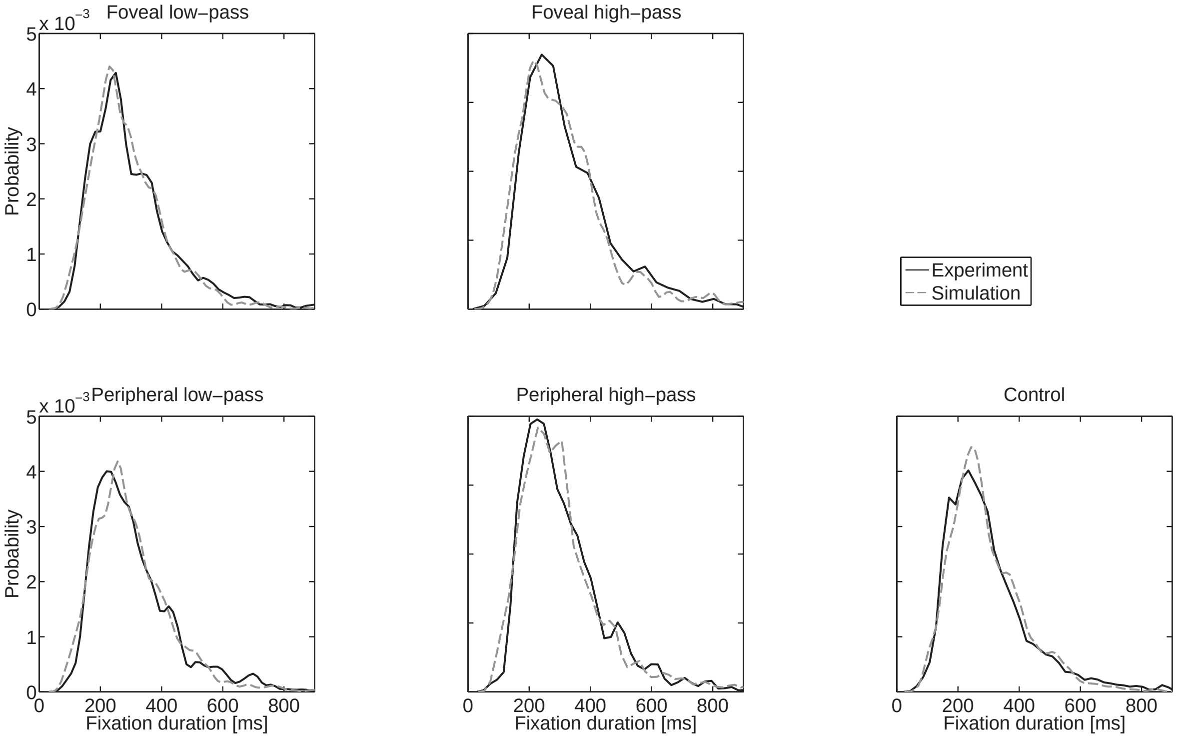

Figure 2.7: Distributions of fixation durations for the experimental data (solid black lines) and the simulated data (dashed gray lines) for one subject. Simulated fixation durations were generated with the set of individual model parameters that resulted from fitting the model to the experimental distributions of this particular subject. Quantile maximum likelihood served as a quantitative goodness-of-fit measure for fitting the distributions.

The average values across subjects for the 12 model parameters are listed in Table 2.3. The mean threshold N_T for the saccade timer is much lower than the values obtained for the thresholds N_F and N_P for foveal and peripheral activation, respectively. With a low timer threshold, fixation duration intervals are more variable, so that the corresponding distributions are considerably skewed (for a numerical simulation on this issue, see Trukenbrod & Engbert, 2014, figure 1b). Furthermore, the peripheral and foveal activations mostly remain in their incrementing phase (see Figure 2.9, upper row). The mean duration of the saccade timer ($t_{sac} = 253$ ms) is slightly larger than the mode for

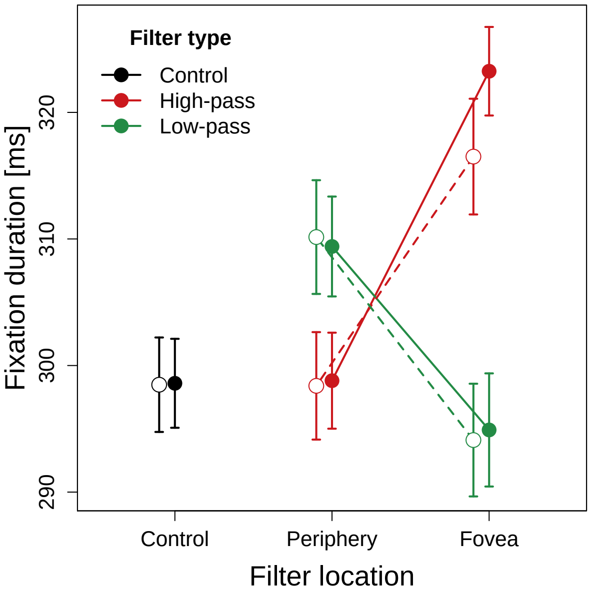

Figure 2.8: Mean fixation durations for the experimental data (filled symbols with solid lines) and the simulated data (open symbols with dashed lines). Error bars represent within-subjects confidence intervals. The simulated data represent the fixation durations for all subjects that were simulated with the 11 sets of individual model parameters.

distributions of fixation durations in scene viewing (Henderson & Hollingworth, 1998). The estimated mean value $\rho = 4.37$ for the strength of foveal inhibition differs significantly from zero, indicating that foveal inhibition is an important model mechanism for explaining the experimental effects. As the numerical values of the parameters j and k indicate, the transition rates for the foveal and peripheral random walks are always higher than the rate for the saccade timer, suggesting that foveal and peripheral activations build up faster than the timer activation. This effect is partly counteracted by the higher thresholds for the foveal and peripheral random walks, so that inhibition can potentially contribute during a substantial part of the fixation duration.

The differing fixation duration effects for the five filter conditions emerge from a modulation of the foveal and peripheral transition rates by variation of the parameters j and k. Table 2.3 indicates that the rates for foveal low-pass and peripheral high-pass filtering, j_{fLP} and k_{pHP}, are similar to the corresponding rates j_C and k_C for the control condition. For foveal high-pass and peripheral low-pass filtering, however, transition rates j_{fHP} and k_{pLP} change drastically compared with the control condition. The foveal transition rate for foveal high-pass filtering increases compared with the foveal rate in the control condition ($j_{fHP} = 3.44$ vs. $j_C = 2.61$), indicating that foveal activation accumulates faster, leading to stronger inhibition. With peripheral low-pass filtering, the peripheral transition rate slows down compared with the corresponding rate for the control condition ($k_{pLP} = 2.53$ vs. $k_C = 3.15$), indicating that peripheral activation accumulates slower, leading to less disinhibition. Both mechanisms increase the activation difference and thus effectively prolong fixation durations by foveal inhibition. Since the increase of fixation durations is not as pronounced for peripheral low-pass filtering as it is for foveal high-pass filtering, the difference between the transition rates (control vs. experimental) is smaller for peripheral low-pass filtering.

Figure 2.9 shows the result of 500 model runs per condition, using the fitted parameter values. It is evident that inhibition is on average active earlier and for a longer time period in the foveal-high pass and the peripheral low-pass than in the other three conditions. However, inhibition plays a significant role in all conditions, especially in generating long fixations. The fraction of time during which inhibition was on is estimated at 43%, 43%, 57%, 53%, and 45% for the control, foveal low-pass, foveal high-pass, peripheral low-pass, and peripheral high-pass conditions, respectively.

2.3.7 Discussion

We developed a model for the control of fixation durations in scene viewing based on the interaction between foveal and peripheral information processing. Numerical simulations of the model reproduced empirical means and distributions of fixation durations well. The model simulations demonstrate that the concept of a dynamical interaction of foveal and peripheral processing is a promising framework for the control of fixation durations.

Specifically, for unimpaired scene viewing (control condition), the model parameters revealed a lower threshold and a higher transition rate of activations in the peripheral compartment compared to the corresponding values in the foveal compartment (see Table 2.3). Therefore, spatially distributed processing in the model is compatible with the current view that, due to the physiological properties of the visual system, processing of information from the peripheral visual field is faster than processing of information from the foveal visual field. In the model, faster rise and decay of peripheral processing leads to a longer time period during which foveal inhibition is active. Therefore, the parameter estimates suggest that fixation duration during normal scene viewing is often controlled by foveal vision.

Model simulations revealed that spatial frequency filtering affected the transition rates for the foveal and peripheral random walks. More precisely, foveal high-pass filtering affected the transition rate for the foveal compartment in the model, and peripheral low-pass filtering affected the transition rate of the peripheral compartment. Consequently, when the model dynamics change with differing transition rates between the five conditions, the amount of foveal inhibition in these conditions changes as well. Since foveal inhibition was active for a substantial proportion of time in each condition, it is an important mechanism for delaying ongoing fixations in each condition.

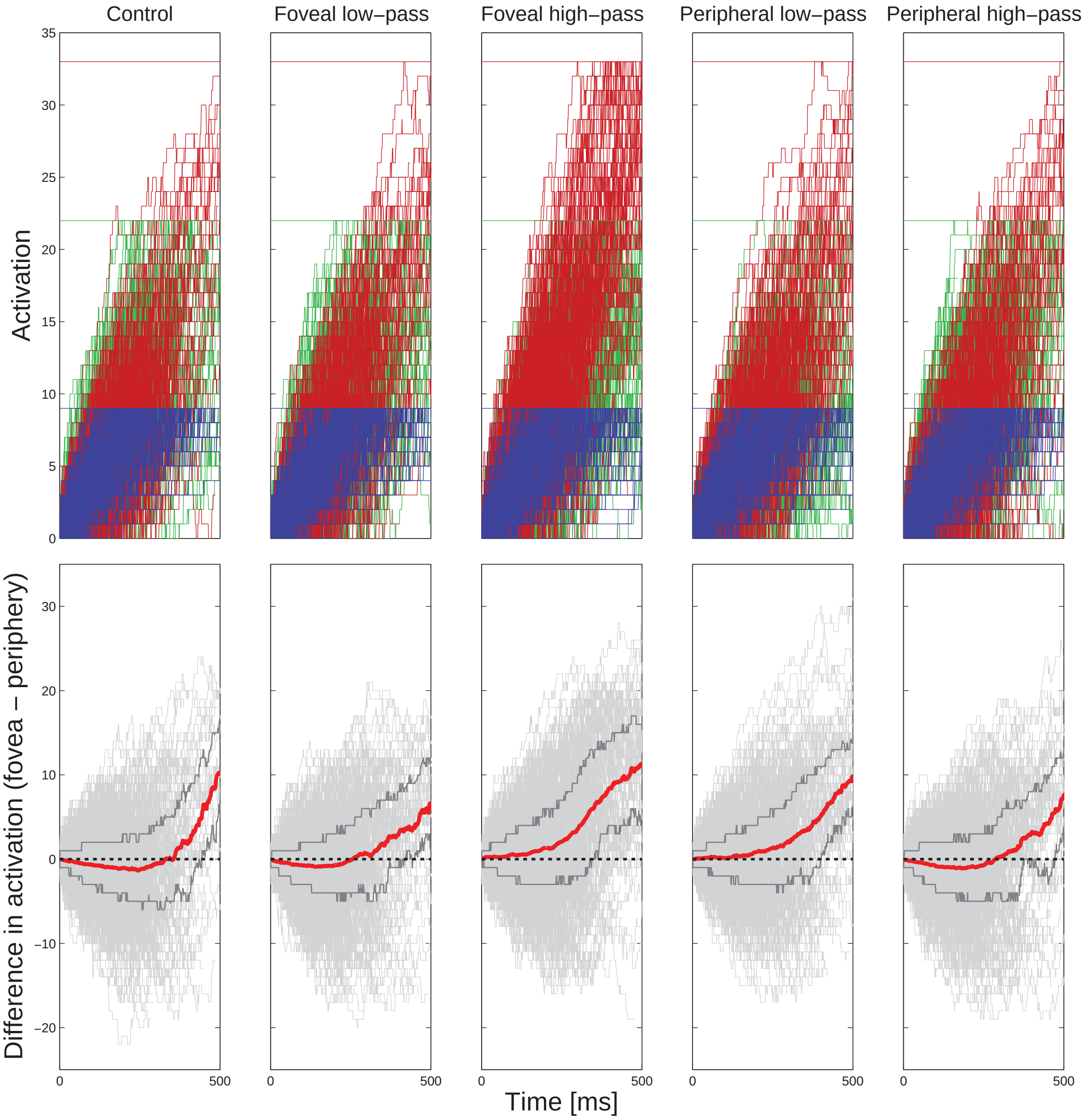

Figure 2.9: Results of 500 model runs per condition, using the fitted parameter values. (Upper row). Evolutions of peripheral (green), foveal (red) and timer (blue) activations over time for control, foveal low-pass, foveal high-pass, peripheral low-pass, and peripheral high-pass conditions (left to right), with the time axis limited to the first 500 ms. Respective thresholds are marked by horizontal lines. (Lower row). Differences of foveal–peripheral activations per condition over time. The thick red line is the average difference, dark gray lines mark the first and third quartile, and the black horizontal dashed line marks a difference of zero. Positive values indicate periods during which inhibition is active. In the control condition and also in foveal low-pass and peripheral high-pass, inhibition is initially suppressed by fast rise of peripheral activations. In contrast, in the peripheral low-pass and especially the foveal high-pass conditions, inhibition on average starts earlier and lasts longer. For short fixations, inhibition does not play a major role except in the latter two conditions. Long fixations tend to be such because of the contribution of inhibition.

Although inhibition is important in all conditions, it makes a stronger contribution in conditions that evoked a larger number of long fixation durations. The model simulations reflect this process in the change of the transition rates. For filter conditions that did not increase mean fixation durations in comparison to the control condition, i.e., foveal low-pass and peripheral high-pass filtering, the model's transition rates were similar to the corresponding rates of the control condition. Thus, the model reproduced the three distributions with unchanged parameter values across conditions and adopted a similar mechanism for saccade timing. Consequently, the mechanism of foveal inhibition was equally important with foveal low-pass filtering, peripheral high-pass filtering, and the control condition. This indicates that the attenuation of critical spatial frequency information did not induce a stronger activation in the fixate center (i.e., higher foveal activation) to take more time for processing the filtered regions, but that the remaining information was not useful enough to make the investment of additional processing time worthwhile.

With foveal high-pass and peripheral low-pass filtering, however, where fixation durations increased significantly compared to the other conditions, the transition rates of the random walks changed markedly. Compared with the rate of the control condition, the foveal transition rate for foveal high-pass filtering increased and was even higher than the peripheral transition rate; for peripheral low-pass filtering, the peripheral transition rate decreased compared to the control condition. Thus, the model adopts two different mechanisms to prolong fixation durations by means of foveal inhibition. With a high foveal transition rate, activation in the foveal compartment builds up faster than activation in the peripheral compartment; with a low peripheral transition rate, peripheral activation builds up slower than foveal activation. Both processes increase the difference between activations in the foveal and peripheral compartments (see Equation 2.3), thus strengthening the foveal inhibition process that slows down the growth rate of the saccade timer and delays the end of the ongoing fixation. The model dynamics for foveal high-pass and peripheral low-pass filtering are essentially different from the model dynamics in the other three conditions. Since foveal high-pass filtering still preserves detailed information that the fovea is specialized in processing, although attenuating other information, fixation durations are prolonged, i.e., more processing time is taken. The model thus induces a stronger activation in the fixate center (i.e., higher foveal activation). Since peripheral low-pass filtering, on the other hand, still preserves useful coarse information for saccade target selection in the periphery (although strongly attenuating edge information that might be useful for separating objects from background), additional processing time is taken in this condition as well. The model reflects this process by reducing activation in the move center.

From these results, we can derive a first interpretation of the relation between time-course of activation and processing difficulty. In the case of high processing difficulty, the saccade timer needs to be inhibited, i.e., the ongoing fixation is prolonged—at least as long as an investment of additional processing time seems worthwhile. In the model, such a prolongation is achieved by a fast build-up of activation in the foveal compartment, i.e., a high transition rate, or by a slow build-up of activation in the peripheral compartment, i.e., a low transition rate. Because of the push-pull type interaction between peripheral and foveal compartments, a fast build-up of activation in the peripheral compartment can compensate the foveal activation and cancel the inhibitory influence from the fovea (disinhibition).

2.4 General discussion

In the present study we investigated the temporal control of eye movements during scene viewing using an experimental manipulation of the spatial frequency content in foveal and peripheral vision and computational modeling. This approach was motivated by the functional segregation of foveal and peripheral vision for high and low spatial frequencies (Banks et al., 1991; Hilz & Cavonius, 1974; Robson & Graham, 1981), and by the lack of knowledge about the foveal and peripheral contributions to the control of fixation durations in scene viewing (but see Trukenbrod & Engbert, 2012).

In the experiment, human observers viewed natural scenes in preparation for a memory test, while high or low spatial frequencies were gaze-contingently filtered in the foveal or peripheral visual field. Following the common notion that fixation durations lengthen with increased processing difficulty (Henderson, 2003; Rayner, 2009), we expected fixation durations to increase with spatial filtering in any case compared with the unfiltered control condition. Fixations were also expected to increase more strongly when filters attenuated the spatial frequency bands that the region of the visual field is best suited to process (fovea: high spatial frequencies, periphery: low spatial frequencies).

Our experimental findings concerning fixation durations turned out to be in partial opposition to the hypotheses. Foveal high-pass filtering and peripheral low-pass filtering caused longer fixation durations compared with the control condition, but foveal low-pass filtering and peripheral high-pass filtering did not affect fixation durations. Thus, with spatial filters that were assumed to impair information processing the most, the temporal control of eye movements was similar to the control condition. Filter conditions where processing should have been relatively easier because the more useful spatial frequency information was still mostly available did increase fixation durations significantly, however. For saccade amplitudes, we replicated effects obtained in previous studies. Peripheral filtering caused a tunnel-vision effect with a preference for saccade targets inside the tunnel (replicating Foulsham et al., 2011; Loschky et al., 2005; Shioiri & Ikeda, 1989), and foveal filtering caused a preference for saccade targets outside the central scotoma (replicating van Diepen, 2001).

The experimental effects of spatial frequency filtering on fixation durations can neither be explained by the different sensitivity of the foveal and peripheral visual field to certain spatial frequency bands, nor by the predicted increase of fixation durations with higher (foveal) processing difficulty. As a solution, we developed a new computational model based on more complicated, dynamical interactions of foveal and peripheral processing for the control of fixation durations. The model was implemented via temporal activations for (a) a random saccade timer, which generates saccadic commands, and (b) foveal and peripheral compartments, which represent the unfolding of foveal and peripheral processing over time. The interaction of foveal and peripheral activations can modulate the saccade timer by foveal inhibition: Higher foveal compared to peripheral activations reflect a bias for inhibiting the saccade timer and prolonging the current fixation; higher peripheral activations, on the other hand, reflect a bias for disinhibiting the saccade timer. This interaction of activations in the foveal and peripheral compartment resembles the balance between the fixate center and the move center in the model framework by Findlay and Walker (1999).

The simulations demonstrated that the interactive model reproduced both the experimental distributions and the mean values for the fixation durations by changing the foveal and peripheral transition rates between the five filter conditions. The transition rates for foveal low-pass filtering and peripheral high-pass filtering did not differ markedly from the rates of the control condition, indicating a similar mechanism for saccade timing. Thus, foveal inhibition was equally strong in these three conditions, suggesting that critical information was so heavily impaired with foveal low-pass and peripheral high-pass filtering that an investment of more processing time in terms of stronger inhibition of the saccade timer did not seem useful. With foveal high-pass filtering and peripheral low-pass filtering, however, the transition rates differed significantly from the rates of the control condition. With foveal high-pass filtering, foveal processing tended to evolve faster and early after the beginning of a fixation, causing a strong inhibition of the saccade timer. High spatial frequencies that the fovea is specialized in processing were still largely preserved, so that the prolongation of fixations by increasing activation in the fixate center was useful to extract as much information as possible. With peripheral low-pass filtering, peripheral processing tended to evolve slower, which also allows a stronger inhibition of the timer. Since filtering impaired processing of peripheral information, but still preserved useful low spatial frequencies, additional processing time was taken by decreasing activation in the move center. Thus, the model generated increased mean fixation durations with both foveal high-pass and peripheral low-pass filtering by a pronounced inhibition of the saccade timer, but the dynamics that produced this behavior were completely different.

Based on a small set of dynamical rules, the model provided a very good fit to the empirical data. In its current version, however, it also has some limitations. First, although the division into two compartments is compatible with and inspired by the two-streams hypothesis (Goodale & Milner, 1992), we do not yet model how the relative weight of the foveal and peripheral compartment is adjusted. It would be interesting to add a mechanism that controls the transitions rates and thresholds for foveal and peripheral processing. Implementation of such a mechanism, which should also be sensitive to different filter characteristics, is left for future work. Second, the model is limited to temporal aspects of eye movement control. Although a full model of eye movement behavior during scene viewing certainly needs to include a mechanism for spatial target selection, we constrained the first version of our computational framework to temporal control in order to obtain a simpler model that might be more easily understood and analyzed. This decision is also justified by the fact that “where” and “when” decisions in the oculomotor system are largely independent (Findlay & Walker, 1999). There is no reason, however, why our model cannot be integrated with one of the many existing models of spatial selection.

We consider our model a fruitful first step at addressing the contribution of different areas of the visual field to the control of fixation duration during scene viewing. Specifically, both experimental data and computational simulations indicate that not only foveal, but also peripheral vision plays a critical role in regulating fixation duration during scene viewing. While current theoretical models of saccade timing suggest that the large variance of fixation durations is due to the decision of where to move the eyes next (Reddi & Carpenter, 2000), our model proposes that the amount of variance in fixation durations might be partially due to the complicated interaction of foveal and peripheral information processing. The model moves beyond the widespread notion that in the time course of a fixation, peripheral information is usually processed to select the next saccade target after the foveal stimulus has largely been analyzed (Rayner, 2009; van Diepen & d’Ydewalle, 2003). Future experimental and computational studies may delineate the general mechanisms that underlie the interaction between foveal and peripheral processing and the underlying adaptivity for a variety of visual tasks.

2.5 Acknowledgments

This work was funded by Deutsche Forschungsgemeinschaft (DFG) by grants LA 2884/1 to J. L. and EN 471/10 to R. E. We thank Petra Schienmann for help with data collection. We are grateful to Casimir Ludwig and an anonymous reviewer for helpful comments.

2.6 Appendix

2.6.1 Estimation of the model parameters

A genetic algorithm (Goldberg, 1989; Holland, 1975) was used to estimate the model parameters. We started with 50 combinations (i.e., individuals) of random parameter values selected from a predefined range that was restricted by mathematical constraints or by neurophysiological plausibility (see Table 2.3). Using selection, mutation, and crossover, individuals were evolved over several thousand generations. After convergence of the population, we used the best 50 individuals of the population from each iteration.

2.6.2 Quantile maximum likelihood

Quantile maximum likelihood estimation (Heathcote et al., 2002) was used as a fitting procedure for the fixation duration distributions. Every experimental distribution was divided into 20 quantiles that each covered 5% of the data. For each parameter combination, simulated fixations for the five distributions were tested for their probability to fall in each of the experimental quantile boundaries (a probability of 0.05 was considered to be a perfect fit). The log likelihood of the data given the parameter values was then computed from the logarithmized probabilities for fixations to fall in each of the 20 quantiles. The 50 individuals with the largest log likelihood were chosen from the population. The source code for the computational model is available at the Potsdam Mind Research Repository (PMR2), http://read.psych.uni-potsdam.de/pmr2.

Chapter 3

Coupling of attention and saccades when viewing scenes with central and peripheral degradation

Anke Cajar, Paul Schneeweiß, Ralf Engbert, and Jochen Laubrock
University of Potsdam, Germany

Running head: Coupling of attention and saccades

---

published 2016 in *Journal of Vision, 16*(2):8, 1–19 in the
Special Issue “Scene perception from central to peripheral vision”,
doi:10.1167/16.2.8.

The online supplements to this article are provided at
http://jov.arvojournals.org/article.aspx?articleid=2513624

Abstract

Degrading real-world scenes in the central or the peripheral visual field yields a characteristic pattern: Mean saccade amplitudes increase with central and decrease with peripheral degradation. Does this pattern reflect corresponding modulations of selective attention? If so, the observed saccade amplitude pattern should reflect more focused attention in the central region with peripheral degradation and an attentional bias toward the periphery with central degradation. To investigate this hypothesis, we measured the detectability of peripheral (Experiment 1) or central targets (Experiment 2) during scene viewing when low or high spatial frequencies were gaze-contingently filtered in the central or the peripheral visual field. Relative to an unfiltered control condition, peripheral filtering induced a decrease of the detection probability for peripheral, but not for central targets (tunnel vision). Central filtering decreased the detectability of central but not of peripheral targets. Additional post hoc analyses are compatible with the interpretation that saccade amplitudes and direction are computed in partial independence. Our experimental results indicate that task-induced modulations of saccade amplitudes reflect attentional modulations.

3.1 Introduction

Visual acuity and spatial resolution are highest at the point of gaze, the fovea, and fall off dramatically to the periphery (see Strasburger, Rentschler, & Jüttner, 2011). Consequently, we have to move our eyes in a sequence of high-velocity saccadic eye movements with intermittent fixations on regions of interest for the exploration of visual detail. During each fixation, the foveal region is analyzed and the next saccade target is selected among competing peripheral locations. More precisely, the visual field is typically divided into three regions: the foveal, the parafoveal, and the peripheral regions. Foveal vision extends to 1° retinal eccentricity and parafoveal vision extends from 1° to 5° eccentricity; the union of both regions is usually referred to as central vision. Peripheral vision comprises the remaining part of the visual field beyond 5° eccentricity (see Larson & Loschky, 2009). How central and peripheral vision interact represents a key problem in scene perception research.

Gaze-contingent filtering of visual information is a powerful experimental tool for investigating the specific contributions of central and peripheral vision to the processing of natural scenes. The moving-window technique, first introduced in reading research (McConkie & Rayner, 1975; Rayner & Bertera, 1979), aligns a window of arbitrary size to the current gaze position of the viewer. In general, the window is centered at the point of gaze and moves with the gaze position in real time. Visual information inside or outside the window can then be altered to manipulate vision in the central or the peripheral visual field, respectively.

Previous research on gaze-contingent scene degradation consistently shows characteristic patterns of mean saccade amplitudes with central versus peripheral filtering. Saccade amplitudes decrease with peripheral filtering (Foulsham et al., 2011; Laubrock, Cajar, & Engbert, 2013; Loschky & McConkie, 2002; Loschky et al., 2005; Nuthmann, 2013, 2014; Nuthmann & Malcolm, 2016; Reingold & Loschky, 2002; Shioiri & Ikeda, 1989; van Diepen & Wampers, 1998) and increase with central filtering (Laubrock et al., 2013; Nuthmann, 2014; Nuthmann & Malcolm, 2016; van Diepen, 2001). These patterns are typically caused by a higher proportion of short saccades with peripheral filtering and a higher proportion of long saccades with central filtering (e.g., see Laubrock et al., 2013). The effects get stronger with increasing filter level (Cajar, Engbert, & Laubrock, 2016; Loschky & McConkie, 2002) and filter size (Cajar, Engbert, & Laubrock, 2016; Loschky & McConkie, 2002; Nuthmann, 2013, 2014). A straightforward interpretation of the results is a viewing strategy of avoiding filtered scene regions as saccade targets. With central filtering, viewers make fewer short inspection saccades in the filtered central region; instead, they program more long saccades that target unfiltered peripheral scene regions. With peripheral filtering, the reverse pattern is obtained: Viewers avoid the filtered periphery and tend to keep their gaze in the unfiltered central region, thus making more short saccades.

These modulations of saccade amplitudes have often been linked to attention (Loschky & McConkie, 2002; Nuthmann, 2013; Reingold & Loschky, 2002); for example, Reingold and Loschky (2002) showed that saccade latencies to peripheral targets in short movie clips and static scenes increase when peripheral information is blurred. However, so far, it has not been directly tested if changes of saccade amplitudes due to varying central or peripheral processing difficulty in scenes indeed reflect corresponding attentional modulations. This is the purpose of the present study.

For the interpretation of eye-movement behavior, the tight coupling between saccades and attention is an important theoretical constraint. Several studies have demonstrated that (a) attention precedes a saccade to an intended target location, and (b) stimulus detection and identification are facilitated at intended saccade target locations compared to other locations in the visual field, at least in simple, reduced task environments (Deubel & Schneider, 1996; Hoffman & Subramaniam, 1995; Kowler et al., 1995). Deubel and Schneider (1996) also showed that it is impossible to discriminate a target at one location while preparing a saccade to another location. Due to this coupling between attention and saccades, saccade amplitudes and direction are thought to reflect attentional selection and thus the spatial extent of parafoveal processing (Nuthmann, 2013; Greene et al., 2010; Loschky & McConkie, 2002). Following this notion, the aforementioned effects of central and peripheral scene degradation on saccade amplitudes suggest more focused attention on the unfiltered central region with peripheral filtering and a stronger attentional bias toward the periphery with central filtering.

However, attention and eye movements can also be partly decoupled as indicated by covert attention shifts. Covert attention shifts can be several times faster than overt eye movements (Nakayama & MacKeben, 1989). Well-known metaphors of attention include the spotlight (Posner, Snyder, & Davidson, 1980), zoom-lens (Eriksen & James, 1986), and the attentional-gradient model (LaBerge & Brown, 1989). In the spotlight model, covert attention moves in an analog fashion across the visual field using disengage, shift, and engage operations. The zoom-lens model extends the spotlight by a variable-sized focus. Several aspects of the spotlight metaphor have been questioned by later work. For example, the movement of covert attention is probably digital rather than analog, meaning that the spotlight is turned off at one location and turned on at the next without passing over intermediate locations (Chastain, 1992a, 1992b; Gersch, Kowler, & Dosher, 2004). Furthermore, the time to move attention between two locations is rather independent of the distance between the two locations (Eriksen & Webb, 1989; Sagi & Julesz, 1985). Although additional effects of object-based selection have been demonstrated in simple tasks (Egly, Driver, & Rafal, 1994; Duncan, 1984) as well as in scene perception (Malcolm & Shomstein, 2015), one critical feature of the spotlight model that has remained valid is that attentional selection is based on location, mirroring the importance of topological maps in the visuospatial processing stream. The attentional gradient model emphasizes that the size and the concentration of attention can vary according to task demands. A further innovative feature is the possibility of several peaks, so that independent locations can be enhanced in parallel (e.g., see Engbert, Trukenbrod, Barthelmé, & Wichmann, 2015, for a computational implementation in scene viewing). This also includes the possibility of a ring-shaped allocation of attention, as has been experimentally demonstrated (Egly & Homa, 1984; Juola, Bouwhuis, Cooper, & Warner, 1991). When the attention field gets large compared to the size of the target, introducing target location uncertainty, one observes a performance decrement with invalid cues (i.e., withdrawal of attention), but no enhancement with valid cues (Herrmann, Montaser-Kouhsari, Carrasco, & Heeger, 2010). In summary, covert attention can obviously be distinguished from overt attention as indicated by eye-movement execution, but it might still be related to or equivalent to eye-movement planning.

To investigate whether modulations of saccade lengths during scene viewing correspond to the spatial extent of attention, we here measured in two experiments how the perceptibility of central or peripheral target objects is affected by gaze-contingent central or peripheral filtering of spatial frequencies. High spatial frequencies carry the fine-grained information of an image whereas low spatial frequencies carry the coarse-grained information of an image. Although the fovea is very sensitive to medium and high spatial frequencies, which are critical for object identification and the analysis of details, the visual periphery is mostly sensitive to low spatial frequencies (Hilz & Cavonius, 1974; Robson & Graham, 1981). Based on this different specialization of central and peripheral vision and on results from a previous study (Laubrock et al., 2013), we assume that high spatial frequencies are more important than low spatial frequencies for processing central information whereas low spatial frequencies are more important than high spatial frequencies for processing peripheral information. We applied low-pass or high-pass filters, which attenuate high or low spatial frequencies, respectively, either to the central or to the peripheral part of the visual field. Such gaze-contingent spatial-frequency filtering systematically alters scene-viewing behavior. In Laubrock et al. (2013), we applied central or peripheral low-pass or high-pass filters to colored scenes, with a gaze-contingent window radius of 2.8°. Replicating the aforementioned effects on saccade amplitudes, amplitudes increased with central and decreased with peripheral filtering compared to an unfiltered control condition. Opposite to our hypotheses, fixation durations increased with central high-pass and peripheral low-pass filtering (the latter replicating Loschky & McConkie, 2002; Loschky et al., 2005; Nuthmann, 2013, 2014; Parkhurst et al., 2000; Shioiri & Ikeda, 1989; van Diepen & Wampers, 1998) but were similar to the unfiltered control with central low-pass and peripheral high-pass filtering. We concluded that fixation durations prolong with increased processing difficulty due to central or peripheral filtering as long as the filter leaves at least a part of the critical information intact (i.e., low spatial frequencies in the peripheral and high spatial frequencies in the central visual field). If most of the critical information is removed, fixation durations can return to the baseline (Laubrock et al., 2013).

In both experiments of the present study, high or low spatial frequencies were filtered either in the central or the peripheral visual field during real-world scene viewing. Participants had the dual tasks of inspecting the scenes carefully in expectation of difficult memory questions and to simultaneously indicate the presence of a green circle in the scene whenever detected. This target stimulus only appeared during randomly chosen fixations in either peripheral (Experiment 1) or central vision (Experiment 2) but always with an offset from the point of fixation. Thus, viewers could only detect the target using covert attention. Consistent with previous findings, we expected mean saccade amplitudes in both experiments to shorten with peripheral filtering and to lengthen with central filtering relative to an unfiltered control condition. If saccadic amplitude modulations are coupled to attention, target detection probability should change in accordance with the effects on saccade amplitudes. Shorter saccade amplitudes with peripheral filtering would therefore reflect tunnel vision (Mackworth, 1965; Williams, 1988) with a shrinkage of the attentional focus, meaning that attention is withdrawn from the filtered periphery. Consequently, the detection of peripheral targets should be impaired whereas the detection of central targets should be unaffected. On the other hand, longer saccade amplitudes with central filtering would reflect a wider attentional focus or more frequent attention shifts to the undegraded periphery as attention is withdrawn from the filtered center. Consequently, the detection of central targets should be impaired with central filtering whereas the detection of peripheral targets should be unaffected.

The hypothesis that central filtering does not impair the detection of peripheral targets challenges previous research showing that an increased foveal load interferes with performance in peripheral detection tasks (Crundall, Underwood, & Chapman, 1999; Holmes et al., 1977; Ikeda & Takeuchi, 1975; Williams, 1985, 1989, 1995). This deterioration in performance has been attributed to a reduction of the functional field of view due to the increased foveal processing demands. Holmes et al. (1977) showed that even the mere presence of a foveal stimulus that subjects were asked to ignore decreases peripheral target detection. In contrast, we hypothesize here that if modulations of saccade amplitudes reflect modulations of attention, peripheral target detection should not deteriorate with central filtering although processing demands in the central visual field are increased. Thus, we predicted perceptual costs in the peripheral detection task with peripheral filtering but not with central filtering.

3.2 Experiment 1

High or low spatial frequencies were filtered either in the central or the peripheral visual field during scene viewing. Viewers explored the scene in preparation for a memory task while simultaneously detecting targets in the visual periphery. Saccade amplitudes were expected to increase with central filtering and to decrease with peripheral filtering relative to an unfiltered control condition with the effects being stronger when filters attenuate more useful spatial frequencies (i.e., central low-pass and peripheral high-pass filtering). If saccade amplitudes reflect attentional selection, peripheral filtering should deteriorate target detection and slow down target response times whereas central filtering should not interfere with detection performance.

3.2.1 Method

3.2.1.1 Participants

Thirty students from the University of Potsdam (four male, mean age: 23 years) with normal or corrected-to-normal vision and normal color discrimination were tested. They received course credit or 12 Euro for participation and were naive as to the purpose of the experiment. The experiment conformed to the Declaration of Helsinki. Participants gave their written informed consent prior to the experiment.

3.2.1.2 Apparatus

Stimuli were presented on a 20-in. Mitsubishi DiamondPro 2070 CRT monitor at a resolution of 1,024 × 768 pixels and a refresh rate of 150 Hz. A head and chin rest ensured stability of the head and a constant viewing distance of 60 cm (23.6 in.). During binocular viewing, gaze position of the right eye was recorded with an EyeLink 1000 tower mount system (SR Research, Ontario, Canada) at a sampling rate of 1000 Hz. Stimulus presentation and response collection were implemented in MATLAB® (The Mathworks, Natick, MA) using the OpenGL-based Psychophysics Toolbox (PTB3; Brainard, 1997; Kleiner et al., 2007; Pelli, 1997) and the Eyelink Toolbox (Cornelissen, Peters, & Palmer, 2002).

3.2.1.3 Stimuli and design

Stimuli were 85 grayscale photographs (1,024 × 768 pixels) of outdoor, real-world scenes that subtended a visual angle of 38.2° × 28.6°. For each scene, low-pass and high-pass filtered versions were prepared in advance via Fourier analysis. Low-pass filtering attenuated spatial frequencies above 1 c/° and high-pass filtering attenuated spatial frequencies below 10 c/°. For gaze-contingent filtering in the central or the peripheral visual field, a foreground and a background image were merged in real time using alpha blending. With peripheral low-pass filtering, for example, the foreground image was the original scene and the background image was the low-pass filtered version of the scene. A 2-D hyperbolic tangent with a slope of 0.06 served as a blending function for creating the alpha mask. The inflection point of the function corresponded to the radius of the gaze-contingent window, which was 5° and thus divided central from peripheral vision. The alpha mask was centered at the current gaze position and defined the transparency value, that is, the weighting of the central foreground image at each point. At the point of fixation, only the foreground image was visible; with increasing eccentricity, the peripheral background image was weighted more strongly until it was fully visible.

Two filter locations (central/peripheral visual field) were crossed with two filter types (low-pass/high-pass filtering), yielding four experimental conditions: central low-pass, central high-pass, peripheral low-pass, and peripheral high-pass filtering. A control condition without filtering served as a baseline (for example stimuli, see Figure 3.1). A Latin square design assured counterbalancing of condition–scene assignments across participants. Scenes were presented in random order.

The target stimulus was a medium-green circle with a diameter of 0.26°; it appeared several times on each trial at a random location 7.5° away from the current gaze position of the viewer. Thus, the circle always appeared in the visual periphery beyond the boundary of the gaze-contingent window and was located on a completely filtered background with peripheral filtering or on a completely unfiltered background with central filtering and the control condition. The target appeared during randomly chosen fixations (every sixth to 10th fixation) about two or three times per scene (mean: 2.8). Online velocity detection in the raw time series of gaze positions was used to identify fixations and saccades during each trial; for a saccade to be detected, the average eye velocity across five consecutive samples (i.e., 5 ms) had to exceed a threshold of 100 °/s, and fixations were designated as periods during which the threshold was not exceeded. Target onset coincided with the beginning of the critical fixation, and target offset coincided with termination of the critical fixation by the viewer or after a maximum presentation time of 300 ms.[1] This means that subjects had to detect the target using peripheral vision (i.e., covert attention) as the target was already removed from the scene when the eyes landed after the next saccade. Target detectability was set to 75% independently for each peripheral scene background (unfiltered, low-pass filtered, high-pass filtered) and individually for each participant using the adaptive QUEST procedure (King-Smith, Grigsby, Vingrys, Benes, & Supowit, 1994; Watson & Pelli, 1983).

A colored circle was chosen as target stimulus because it was independent from the grayscale scene due to its color feature. It is better suited as a target than, for example, a Gabor stimulus, which shares all its features with the scene (spatial frequency, contrast, color, orientation) and therefore interacts strongly with the background; thus, its detectability can range from very easy to very difficult, depending on the scene patch on which it is located. The colored circle promised more consistent visibility across different scene patches. Medium green was chosen over other colors that appeared too salient on the grayscale scenes.

3.2.1.4 Procedure

Data from each participant were collected in two sessions. In the first session, individual transparency thresholds for 75% target detection on unfiltered, low-pass filtered, and high-pass filtered backgrounds were determined. Each participant viewed three blocks of either unfiltered, completely low-pass filtered, or completely high-pass filtered scenes. The eye-tracker was calibrated at the beginning of each block and after every 15 trials. Each trial started with a gray screen showing a central fixation point. If the point had been fixated for at least 50 ms within one second from trial start, the scene was revealed; otherwise, a recalibration was initiated. Each scene was presented for 12 s. Participants were instructed to explore the scene and memorize the scene content. In addition, they were told to indicate the presence of the target stimulus by pressing the computer mouse button as fast as possible. The response deadline for the target was reached after 1.5 seconds; later responses were labeled as misses. The transparency (alpha channel) of the target was adjusted using the QUEST algorithm, a Bayesian adaptive method for measuring psychometric thresholds. The QUEST staircase adjusted the transparency of the target on the scene between 0 and one—smaller alpha values indicate a higher transparency of the target and thereby impede its segregation from the scene background. The individual transparency thresholds determined by QUEST then served

---

[1] More precisely, the target was set on and off at the next possible display refresh, usually 6.7 ms after detection of the event.

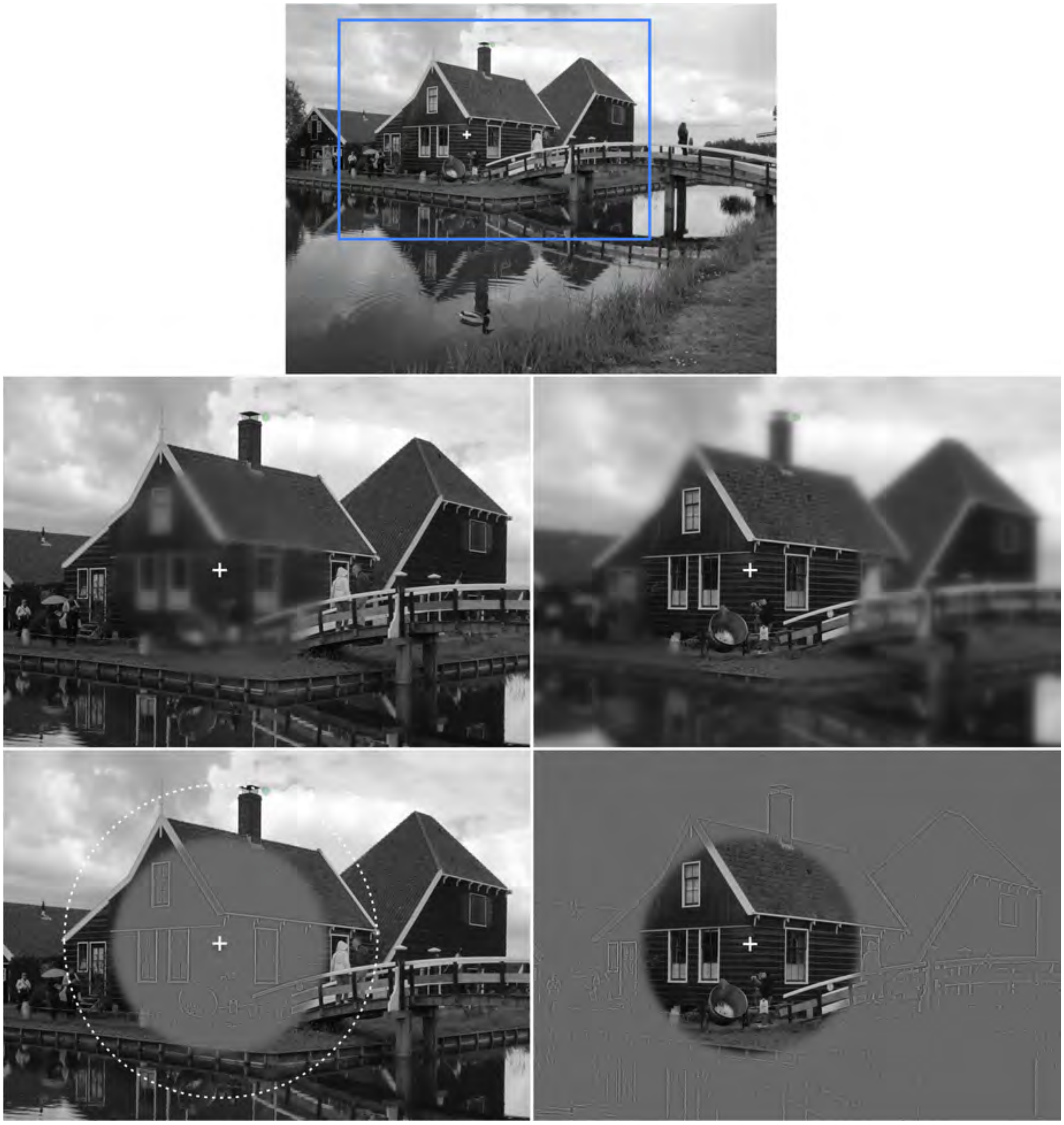

Figure 3.1: Illustration of the five conditions and the peripheral target stimulus in Experiment 1. The white cross indicates the current gaze position, and the dashed-line circle (see lower left image) indicates the possible target positions around the gaze at 7.5° eccentricity. In this example, the green target is located above the gaze. Target visibility varies with unfiltered, low-pass filtered, and high-pass filtered backgrounds because target detectability was set to 75% on each background (see Procedure). The top panel shows the original stimulus in the unfiltered control condition. Below, the four filter conditions are illustrated with cropped and zoomed-in versions of the original stimulus (indicated by the blue frame) to better illustrate the filter effects: (second row, left) central low-pass filtering, (second row, right) peripheral low-pass filtering, (third row, left) central high-pass filtering, (third row, right) peripheral high-pass filtering. Note that these images serve as illustrations; because of re-sampling in reproduction and because the illustrated scenes are at a smaller scale than presented in the experiment, they do not faithfully reproduce the filters actually used in the experiment.

as fixed target transparencies for unfiltered, low-pass filtered, and high-pass filtered peripheral backgrounds in the second session (i.e., the main experiment).

The main experiment presented scenes with gaze-contingent spatial frequency filtering. The eye-tracker was calibrated at the beginning of the experiment and after every 15 trials. Two practice trials were given to acquaint participants with the gaze-contingent display and the tasks. As in the first session, each trial started with a central fixation trigger. Participants viewed 85 scenes that were each presented for 12 s. Varying from trial to trial, scenes were either presented unfiltered or with central or peripheral low-pass or high-pass filtering. For each of the five conditions, 17 scenes were presented. Participants were confronted with two tasks. First, they were instructed to inspect the scenes carefully in preparation for questions about the scene content. Additionally, they were asked to indicate the presence of a green circle as fast as possible whenever detected by pressing the computer mouse button. To ensure that viewers actually engaged in processing the scene and not merely focused on target detection on a scene background, they were asked to answer a three-alternative question about the scene content after a randomly chosen 50% of the scenes. Questions typically asked about the presence or absence of certain objects in the scene (e.g., “Which object was seen behind the car?”), about the location of objects (e.g., “Where in the scene was the Ferris wheel standing?”), or about the number of certain objects (e.g., “How many people were present in the scene?”). The memory question for the scene in Figure 3.1, for example, was “Which object was not present in the scene?” with the response alternatives “duck”, “street lamp”, or “other bird” (correct answer: “street lamp”). As each scene was presented for a long time, scenes were fairly rich in detail, and questions were constructed to be rather difficult to ensure that viewers would carefully explore each scene until the end of the trial. In a previous study (Cajar, Engbert, & Laubrock, 2016) in which we applied central and peripheral spatial-frequency filters to a similar stimulus set, including a majority of the present scenes, viewers answered the memory questions correctly in about 70-75% of the cases although they had no other task. Since viewers in the present experiments were simultaneously confronted with target detection as a second task, we expected somewhat worse memory performance.

3.2.1.5 Data preparation

Saccades were detected in the raw time series of gaze positions using a velocity-based algorithm (Engbert & Kliegl, 2003; Engbert & Mergenthaler, 2006) with a relative velocity threshold of 5 *SD* and a minimum duration of six data samples. A total of 28 trials (1.1%) were removed owing to poor calibration or too much data loss. Single fixations and saccades were removed if they neighbored eye blinks or if they were the first or last event of a trial and therefore associated with scene onset or offset. Overall, 80,647 fixations and 83,762 saccades remained for general eye-movement analyses, and 7,150 valid critical fixations during which the target stimulus was presented were left for analyses.

3.2.1.6 Data analyses

Dependent variables were analyzed using linear mixed-effects models (LMMs) and generalized linear mixed-effects models (GLMMs) as implemented in the *lme4* package (Bates, Maechler, Bolker, & Walker, 2015) of the *R* system for statistical computing (version 3.2.3; R Core Team, 2015). In addition to the fixed effects for experimental conditions, mixed-effects models simultaneously account for random effects due to differences between subjects and items (i.e., scenes). (G)LMMs can account for the variance between subjects and items (a) in the mean of the dependent variable (random intercepts), (b) for all experimental conditions (random slopes), and (c) for correlations between intercepts and slopes. To identify the proper random-effects structure for each model, we applied the algorithm suggested by Bates, Kliegl, Vasishth, and Baayen (2015). For fixed effects, GLMMs yield regression coefficients, standard errors, z values and p values. LMMs only yield regression coefficients, standard errors, and t values because the degrees of freedom are not known exactly for LMMs. For large data sets as in the present experiment, however, the t distribution converges to the standard normal distribution for all practical purposes; consequently, t statistics of an absolute value of 1.96 or larger are considered significant on the two-tailed 5% level (Baayen, Davidson, & Bates, 2008, note 1). All (G)LMMs for Experiment 1 are provided in the online supplements to this article.

3.2.2 Results

3.2.2.1 Eye-movement behavior

Fixation durations and saccade amplitudes were analyzed using LMMs. Fixed effects were estimated using contrast coding that tested for (a) a difference between the control condition and all experimental conditions, (b) a main effect of filter type (low-pass vs. high-pass), (c) a main effect of filter location (periphery vs. center), and (d) an interaction between filter type and filter location. As distributions of fixation durations and saccade amplitudes were positively skewed, variables were transformed before model fitting to approximate normally distributed model residuals. To find a suitable transformation, the optimal λ-coefficient for the Box-Cox power tranformation (Box & Cox, 1964) was estimated using the *boxcox* function of the *MASS* package (Venables &

Ripley, 2002) with $y(\lambda) = (y^{\lambda} - 1)/\lambda$ if $\lambda \neq 0$ and $log(y)$ if $\lambda = 0$. For fixation durations, the log-transformation was optimal (λ = -0.10), and for saccade amplitudes, the transformation with $\lambda = 0.30$ was optimal. Critical fixations (during which the target was presented) and saccades that followed these critical fixations were excluded from the analyses because they likely reflect different eye-movement behavior in response to the target stimulus.

Saccade amplitudes Mean saccade amplitudes are illustrated in Figure 3.2a. As expected from previous research, mean amplitudes lengthened with central and shortened with peripheral filtering relative to the unfiltered control condition. This is indicated by a strong main effect of filter region ($b = 1.78 \cdot 10^{-1}, SE = 1.07 \cdot 10^{-2}, t = 16.66$). Thus, viewers preferred unfiltered scene regions as saccade targets. The effect is also reflected in the distributions of saccade amplitudes (Figure 3.2c); with peripheral filtering, viewers placed more short saccades in the unfiltered central region and fewer long saccades in the filtered periphery compared with the unfiltered control. With central filtering, on the other hand, a higher proportion of long saccades and a lower proportion of short saccades was observed. The effects were stronger when critical spatial frequencies in the central or the peripheral visual field were missing (i.e., with central low-pass filtering and peripheral high-pass filtering). Low-pass filtering therefore caused longer saccades than high-pass filtering with both filter locations, which is indicated by a main effect of filter type ($b = 5.36 \cdot 10^{-2}, SE = 4.67 \cdot 10^{-3}, t = 11.49$). For foveal analysis, low spatial frequencies are less important than high spatial frequencies; this led to fewer short inspection saccades with central low-pass filtering than with central high-pass filtering. High spatial frequencies, on the other hand, are of little use for peripheral target selection. Viewers therefore targeted the filtered periphery less often with peripheral high-pass filtering than with low-pass filtering and instead placed a higher proportion of short saccades in the unfiltered central region. The LMM also revealed an interaction between filter type and filter region ($b = 2.34 \cdot 10^{-2}, SE = 5.98 \cdot 10^{-3}, t = 3.92$) as the difference between filter types was slightly greater with central filtering than with peripheral filtering.

Fixation durations Mean fixation durations are illustrated in Figure 3.2b. Fixation duration averaged across the four filter conditions significantly increased relative to unfiltered scene viewing ($b = 3.51 \cdot 10^{-2}, SE = 7.62 \cdot 10^{-3}, t = 4.60$). Mean fixation duration was longer with central filtering than with peripheral filtering ($b = 5.64 \cdot 10^{-2}, SE = 9.20 \cdot 10^{-3}, t = 6.13$). There was no main effect of filter type ($b = 1.19 \cdot 10^{-2}, SE = 8.89 \cdot 10^{-3}, t = 1.34$). However, an interaction between filter type and filter

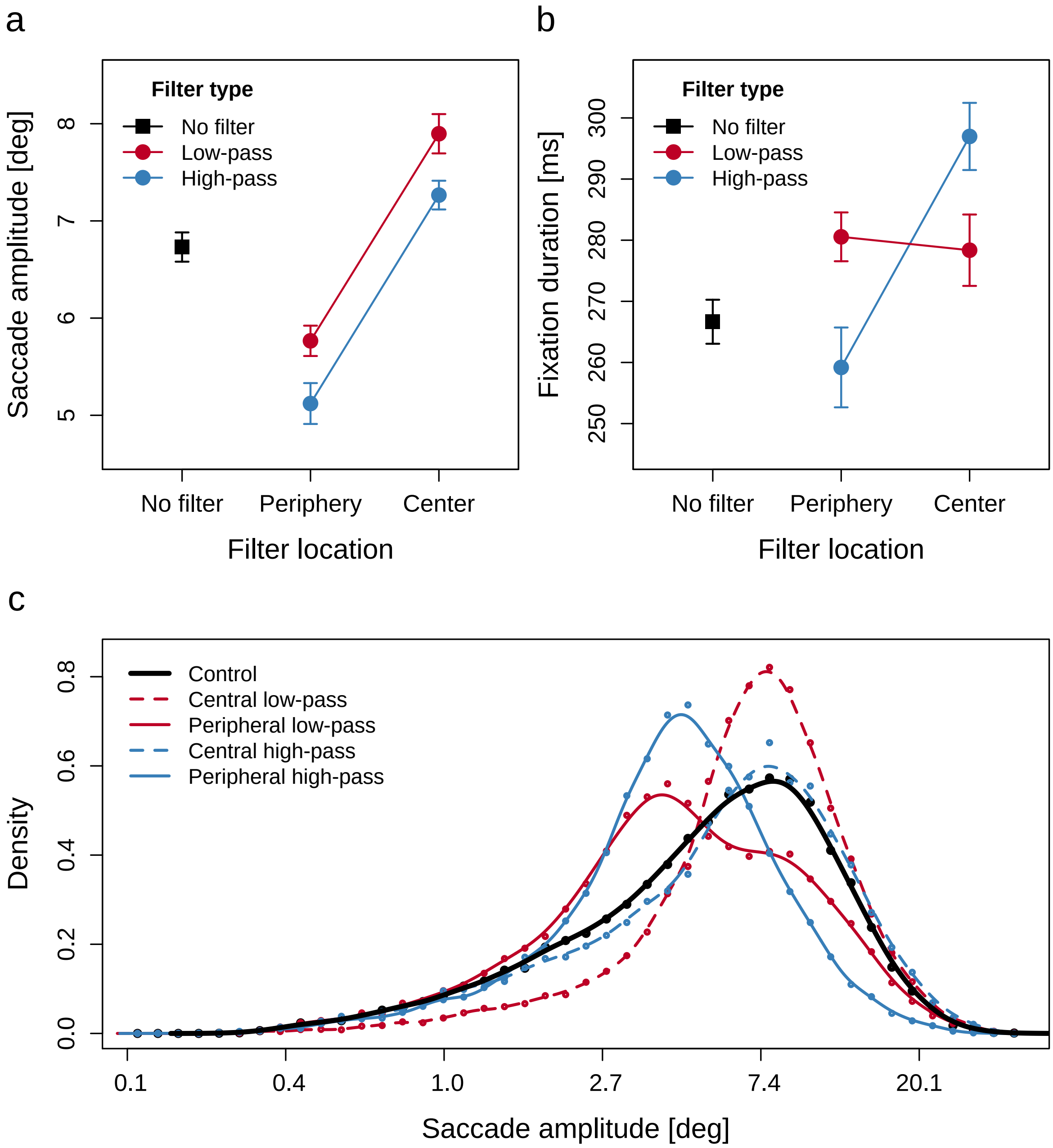

Figure 3.2: Eye-movement behavior in Experiment 1. Mean saccade amplitudes (a) and mean fixation durations (b). Error bars are 95% within-subject confidence intervals with the Cousineau-Morey correction applied (Cousineau, 2005; Morey, 2008). (c) Distributions of saccade amplitudes. The logarithmic scaling of the abscissa emphasizes the effects for short saccades. Lines represent kernel density estimates using a Gaussian kernel as implemented in the *R* function *density*; the bandwidth was chosen according to Silverman, 1986, equation 3.31 with a weight of 1.2. The area under each curve adds up to one.

region occurred ($b = -7.17 \cdot 10^{-2}, SE = 8.58 \cdot 10^{-3}, t = -8.36$). With central filtering fixation durations were longer with high-pass filtering than with low-pass filtering, but with peripheral filtering fixation durations were longer with low-pass filtering than with high-pass filtering. Mean fixation duration even numerically dropped below the baseline with peripheral high-pass filtering (see Figure 3.2b). The effects replicate our results from a previous study (Laubrock et al., 2013) and indicate that fixations were prolonged more markedly when critical spatial frequencies were still available (center: high frequencies; periphery: low frequencies).

3.2.2.2 Task performance

Correct answer probability for the memory questions and target detection probability were analyzed using GLMMs with a logit link function. Reaction times to the target were analyzed with an LMM. Because the reaction time distribution was positively skewed, reaction times were transformed with λ = -0.42. Reaction times shorter than 260 ms ($N = 25$) and longer than 1200 ms ($N = 25$) were identified as clear outliers based on the raw reaction time distribution and were therefore excluded from the analysis. Fixed effects for all three dependent variables were estimated with treatment contrasts as implemented in *R* that tested for differences between each filter condition and the unfiltered control condition.

Memory questions Mean proportions of correct answers to memory questions about the scene content were 69.6% for the unfiltered control condition, 59.1% for central low-pass filtering, 66.0% for central high-pass filtering, 65.1% for peripheral low-pass filtering, and 63.4% for peripheral high-pass filtering. These values are far above the guessing probability and not far below the proportion of 70%-75% correct that we observed in a previous study (Cajar, Engbert, & Laubrock, 2016) using the same stimuli and scene encoding as a single task. We are therefore assured that viewers in the present experiment engaged in scene processing and did not disregard scene encoding in favor of target detection. The GLMM showed that only central low-pass filtering significantly differed from the control condition in the proportion of correct answers ($b = -0.45, SE = 0.21, z = -2.16, p = 0.031$).

Target detection Transparency thresholds for 75% target detection across participants were 0.36 ($SD = 0.07$) for unfiltered backgrounds, 0.24 ($SD = 0.07$) for low-pass filtered backgrounds, and 0.17 ($SD = 0.04$) for high-pass filtered backgrounds; the target was therefore most transparent on high-pass filtered backgrounds and least transparent on unfiltered backgrounds with low-pass filtered backgrounds in between.

Overall target detection probability was 72%. Mean probabilities for the five conditions are shown in Figure 3.3a. Target detection probability with central filtering did not differ from the control condition for either filter type ($b = -9.15 \cdot 10^{-2}, SE = 1.02 \cdot 10^{-1}, z = -0.90, p = 0.369$ for central low-pass filtering and $b = -4.21 \cdot 10^{-2}, SE = 1.12 \cdot 10^{-1}, z = -0.38, p = 0.707$ for central high-pass filtering). Peripheral filtering, on the other hand, decreased target detection probability relative to the control condition ($b = -3.09 \cdot 10^{-1}, SE = 9.08 \cdot 10^{-2}, z = -3.40, p < 0.001$ for peripheral low-pass filtering and $b = -8.09 \cdot 10^{-1}, SE = 1.83 \cdot 10^{-1}, z = -4.42, p < 0.001$ for peripheral high-pass filtering).

Overall reaction time to the target was 576 ms. Mean reaction times to the target for the five conditions are shown in Figure 3.3b. With central filtering, reaction times did not differ from the control condition for either filter type ($b = -2.49 \cdot 10^{-4}, SE = 2.62 \cdot 10^{-4}, t = -0.95$ for central low-pass filtering and $b = -3.52 \cdot 10^{-4}, SE = 3.09 \cdot 10^{-4}, t = -1.14$ for central high-pass filtering). Reaction times slowed down, however, with both peripheral filters ($b = -8.17 \cdot 10^{-4}, SE = 2.78 \cdot 10^{-4}, t = -2.94$ for peripheral low-pass filtering and $b = -2.23 \cdot 10^{-3}, SE = 3.96 \cdot 10^{-4}, t = -5.62$ for peripheral high-pass filtering).

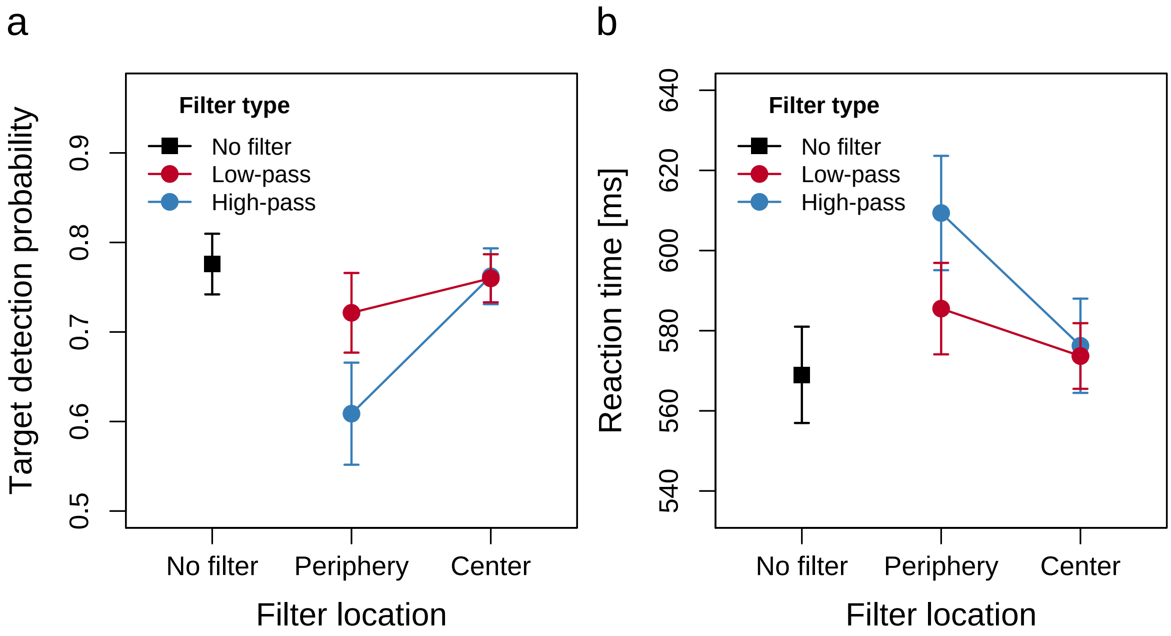

Figure 3.3: Task performance in Experiment 1. Mean target detection probability (a) and mean reaction time to the target (b). Error bars are 95% within-subject confidence intervals with the Cousineau-Morey correction applied (Cousineau, 2005; Morey, 2008).

As participants were allowed to move their eyes freely during scene viewing, we were also interested in the coupling between eye movements and the distribution of covert attention. We therefore investigated how target detection, which could only be achieved using peripheral vision, was correlated with the size and direction of the saccade that followed the critical fixation. For this purpose, target detection probability was analyzed as a function of the angle between target position and landing position as well as amplitude of the subsequent saccade (see Figure 3.4a for further details). A saccade that landed on the former target position corresponds to a target–saccade angle of 0° of arc and a saccade amplitude of 7.5° of visual angle. Figure 3.4b shows a plot of target detection probability across all five conditions as a function of target–saccade angle (in steps of 45° of arc) and saccade amplitude (divided into three bins). Expectedly, target detection probability was highest when the subsequent saccade landed near the former position of the target, supporting the findings that saccades and attention are tightly coupled (Deubel & Schneider, 1996; Hoffman & Subramaniam, 1995; Kowler et al., 1995). Detection probability appeared worst when saccades longer than 10° of visual angle were made that also landed more than 67° of arc away from the previous target position. Most interestingly, target detection probability was generally best when saccade amplitudes were similar to the target's distance from the fovea rather independent of saccade direction[2]. A GLMM using a logit link function and treatment contrasts tested for differences of target detection probability between the three ranges of saccade amplitude ($< 5°$, 5° to 10°, and $> 10°$ of visual angle). The amplitude range from 5° to 10°, which includes the target's distance from the fovea, served as the reference category ($b = 1.61, SE = 0.12, z = 13.56, p < 0.001$). Compared to this amplitude range, target detection probability significantly decreased for saccade amplitudes $< 5°$ ($b = -1.24, SE = 0.12, z = -9.97, p < 0.001$) and for amplitudes $> 10°$ ($b = -1.59, SE = 0.12, z = -13.66, p < 0.001$). These results might indicate a ring-shaped allocation of covert attention induced by the present task demands with which several peripheral locations are scanned in parallel for the detection of the target, of which only one is chosen as the next saccade target.

3.3 Experiment 2

Experiment 1 found support for the hypothesis that reductions of saccade amplitude due to peripheral filtering involve a corresponding withdrawal of attention from the peripheral visual field. Experiment 2 tested whether increased saccade amplitudes with central filtering reflect a withdrawal of attention from the central visual field. As in Experiment 1, high or low spatial frequencies were filtered either in the central or the peripheral visual field during scene viewing. Viewers explored the scene in preparation for a memory task while simultaneously detecting targets. In contrast to Experiment 1, targets were presented in the central rather than the peripheral visual

---

[2] Note that this effect was consistently found for all five filter conditions separately (not shown).

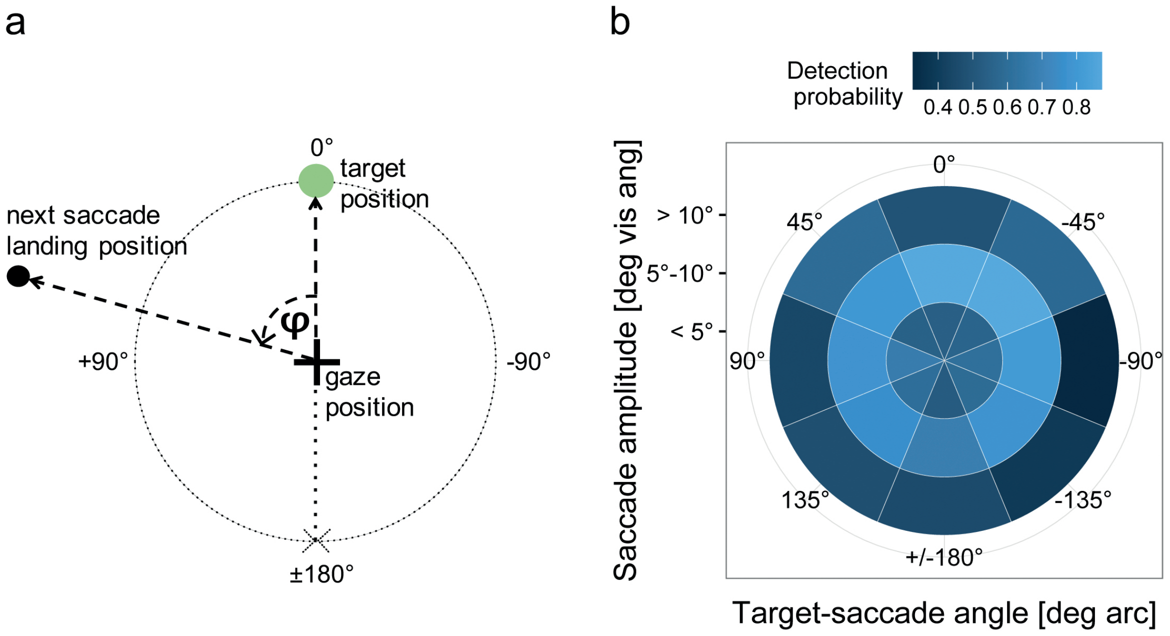

Figure 3.4: Target detection probability as a function of target–saccade angle. (a) The target–saccade angle φ was computed as angle of the arc between vectors connecting the gaze position during the critical fixation with the position of the target and the landing position of the subsequent saccade. An angle of 0° of arc corresponds to a saccade in the direction of the target. Positive angles ranging from 0° to +180° correspond to saccade landing positions to the left of the target whereas negative angles ranging from 0° to -180° correspond to saccade landing positions to the right of the target. (b) Target detection probability as a function of target–saccade angle and saccade amplitude in Experiment 1. Probability was computed for target–saccade angles in steps of 45° of arc and for three ranges of saccade amplitude: $<$ 5°, 5°-10°, and $>$ 10° of visual angle. The medium amplitude range (5°-10°) comprised all possible target positions at 7.5° retinal eccentricity. Lighter colors in the plot correspond to a higher probability to detect the target.

field. Saccade amplitudes were expected to increase with central filtering and to decrease with peripheral filtering relative to the unfiltered control, especially with central low-pass and peripheral high-pass filtering. If saccade amplitudes reflect attention, central filtering should reduce central target detection and slow down target response times whereas peripheral filtering should not interfere with central target detection.

3.3.1 Method

3.3.1.1 Participants

Participants were 30 students from the University of Potsdam (three male, mean age: 22 years) with normal or corrected-to-normal vision and normal color discrimination. None of them participated in Experiment 1. They received course credit or 12 Euro for participation and were naive as to the purpose of the experiment. The experiment conformed to the Declaration of Helsinki. Participants gave their written informed consent prior to the experiment.

3.3.1.2 Apparatus

The apparatus was the same as in Experiment 1.

3.3.1.3 Stimuli and design

Stimuli and design were the same as in Experiment 1. The only difference was that the target was not presented in the peripheral but in the central visual field at a random location 3.5° away from the current fixation point.

3.3.1.4 Procedure

The procedure was the same as in Experiment 1.

3.3.1.5 Data preparation

A total of 27 trials (1.0%) were removed owing to poor calibration or too much data loss. Overall, 75,888 fixations and 79,164 saccades remained for eye-movement analyses. Furthermore, 6,718 valid critical fixations during which the target stimulus was presented were left for analyses.

3.3.1.6 Data analyses

Data analyses were performed as in Experiment 1. All (G)LMMs for Experiment 2 are provided in the online supplements to this article.

3.3.2 Results

3.3.2.1 Eye-movement behavior

LMMs were performed as in Experiment 1. Fixation durations were log-transformed before model fitting, and saccade amplitudes were transformed with $\lambda = 0.26$. Critical fixations during which the target was presented and saccades that followed these critical fixations were excluded from the analyses.

Saccade amplitudes Mean saccade amplitudes are illustrated in Figure 3.5a. Results parallel those of Experiment 1. Mean saccade amplitudes increased with central filtering and decreased with peripheral filtering, leading to a strong main effect of filter location ($b = 1.72 \cdot 10^{-1}, SE = 8.75 \cdot 10^{-3}, t = 19.67$). Furthermore, a main effect of filter type indicated longer amplitudes with low-pass filtering than with high pass filtering with both filter locations ($b = 5.53 \cdot 10^{-2}, SE = 5.99 \cdot 10^{-3}, t = 9.23$). The distributions of saccade amplitudes (see Figure 3.5c) also reflect the same pattern as in Experiment 1. With central filtering, the amount of short saccades decreased whereas the amount of longer saccades increased, particularly with low-pass filtering. With peripheral filtering, on the other hand, the amount of short saccades increased and the amount of long saccades decreased, particularly with high-pass filtering. An interaction of filter type and filter location ($b = 1.86 \cdot 10^{-2}, SE = 7.63 \cdot 10^{-3}, t = 2.44$) occurred because the difference between filter types was greater with central than with peripheral filtering. Note that the distribution of saccade amplitudes in the control condition suggests that participants did not prioritize the target-detection task; otherwise, a pronounced mode of the distribution would be expected at the target eccentricity of 3.5°.

Fixation durations Mean fixation durations are illustrated in Figure 3.5b. Fixation duration averaged across the four filter conditions did not increase relative to the unfiltered control ($b = 1.25 \cdot 10^{-2}, SE = 1.02 \cdot 10^{-2}, t = 1.2$). With peripheral high-pass filtering mean fixation duration was even numerically shorter than in the control condition (see Figure 3.5b). Fixations were significantly longer with central than with peripheral filtering ($b = 5.66 \cdot 10^{-2}, SE = 1.12 \cdot 10^{-2}, t = 5.1$). A main effect of filter type also indicated longer fixation durations with low-pass than with high-pass filtering ($b = 3.54 \cdot 10^{-2}, SE = 9.96 \cdot 10^{-3}, t = 3.6$). However, this main effect was qualified by an interaction between filter type and filter location ($b = -3.84 \cdot 10^{-2}, SE = 7.38 \cdot 10^{-3}, t = -5.2$), showing that fixation durations were indeed longer with low-pass filtering than with high-pass filtering in the periphery but that there was no difference between filter types with central filtering.

3.3.2.2 Task performance

GLMMs and the LMM were performed as in Experiment 1. Reaction times were transformed before model fitting with λ = -0.42. Reaction times shorter than 240 ms (N = 17) were identified as clear outliers based on the raw reaction time distribution and the LMM residuals and were therefore excluded from the analysis.

Memory questions Mean proportions of correct answers to the memory questions about the scene content were 70.8% for the unfiltered control condition, 60.9% for central low-pass filtering, 63.0% for central high-pass filtering, 66.8% for peripheral low-pass filtering, and 66.3% for peripheral high-pass filtering. These values assured us that viewers in the present experiment actually engaged in scene processing and did not

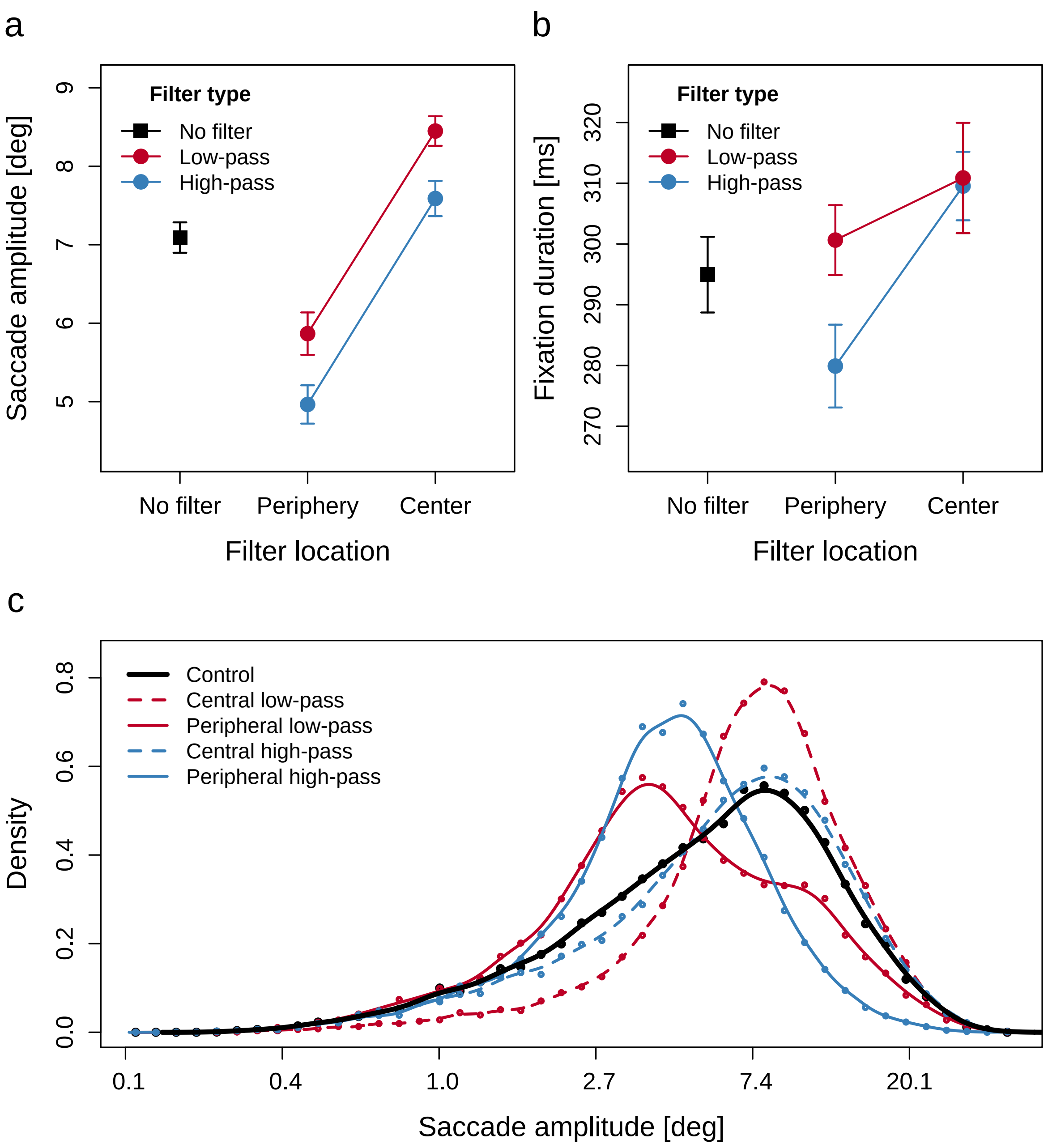

Figure 3.5: Eye-movement behavior in Experiment 2. Mean saccade amplitudes (a) and mean fixation durations (b). Error bars are 95% within-subject confidence intervals with the Cousineau-Morey correction applied (Cousineau, 2005; Morey, 2008). (c) Distributions of saccade amplitudes. The logarithmic scaling of the abscissa emphasizes the effects for short saccades. Lines represent kernel density estimates using a Gaussian kernel as implemented in the *R* function *density*; the bandwidth was chosen according to Silverman, 1986, equation 3.31 with a weight of 1.2. The area under the curve adds up to one.

disregard scene encoding in favor of target detection. The decrease in the proportion of correct answers relative to the control condition was significant with central low-pass filtering ($b = -0.52, SE = 0.21, z = -2.44, p = 0.015$) and marginally significant with central high-pass filtering ($b = -0.41, SE = 0.21, z = -1.91, p = 0.056$). The decrease with peripheral filtering was not significant.

Target detection Transparency thresholds for 75% target detection across participants were 0.21 ($SD = 0.04$) for unfiltered backgrounds, 0.14 ($SD = 0.04$) for low-pass filtered backgrounds, and 0.10 ($SD = 0.02$) for high-pass filtered backgrounds. Thresholds were lower than in Experiment 1, because the target was closer to the fovea and therefore easier to detect.

Overall target detection probability was 67.2%. Mean detection probabilities for the five conditions are shown in Figure 3.6a. Detection probability with peripheral filtering was similar to the unfiltered control condition for both filter types ($b = -1.30 \cdot 10^{-1}, SE = 1.00 \cdot 10^{-1}, z = -1.30, p = 0.195$ for high-pass filtering and $b = 9.01 \cdot 10^{-3}, SE = 9.41 \cdot 10^{-2}, z = 0.10, p = 0.924$ for low-pass filtering). With central filtering, on the other hand, target detection probability significantly decreased with both filter types ($b = -3.70 \cdot 10^{-1}, SE = 8.74 \cdot 10^{-2}, z = -4.23, p < 0.001$ for low-pass filtering and $b = -1.42, SE = 1.75 \cdot 10^{-1}, z = -8.08, p < 0.001$ for high-pass filtering). Contrary to our predictions, this decrease was more dramatic with central high-pass filtering.

Overall reaction time to the target was 619 ms. Mean reaction times for the five conditions are illustrated in Figure 3.6b. Reaction time did not differ from the unfiltered control with peripheral low-pass filtering ($b = -4.09 \cdot 10^{-4}, SE = 2.81 \cdot 10^{-4}, t = -1.45$) but increased with all other filter conditions, especially with central high-pass filtering ($b = -7.35 \cdot 10^{-4}, SE = 3.32 \cdot 10^{-4}, t = -2.21$ for peripheral high-pass filtering; $b = -7.59 \cdot 10^{-4}, SE = 2.77 \cdot 10^{-4}, t = -2.74$ for central low-pass filtering; $b = -2.84 \cdot 10^{-3}, SE = 3.64 \cdot 10^{-4}, t = -7.82$ for central high-pass filtering).

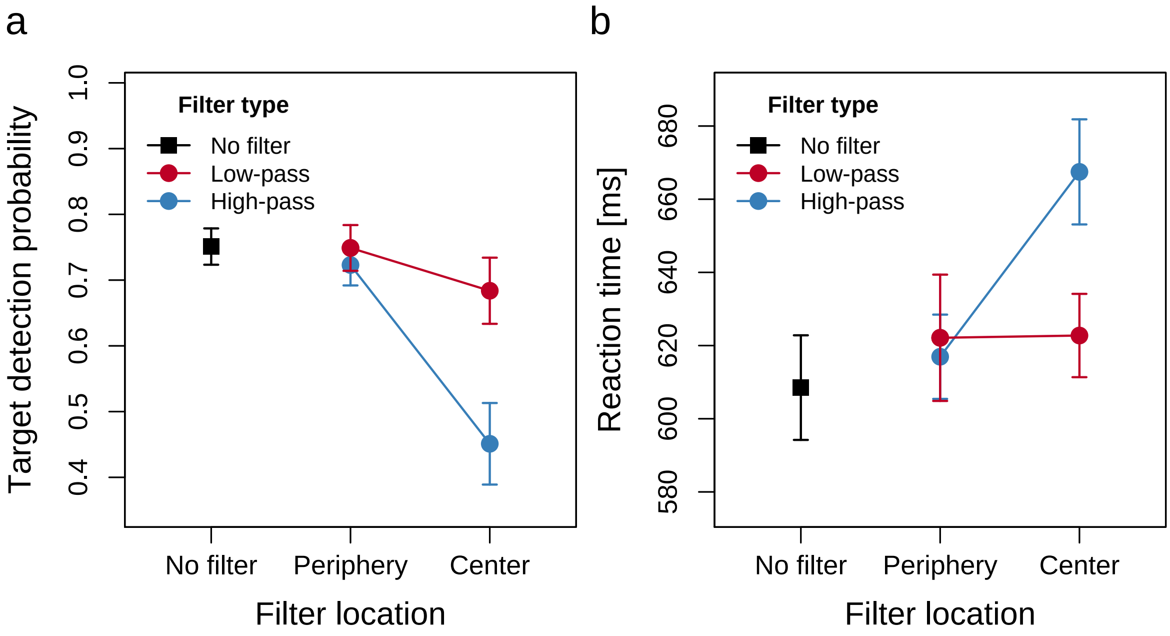

Figure 3.6: Task performance in Experiment 2. Mean target detection probability (a) and mean reaction time to the target (b). Error bars are 95% within-subject confidence intervals with the Cousineau-Morey correction applied (Cousineau, 2005; Morey, 2008).

As in Experiment 1, we investigated how target detection depended on the size and direction of the saccade following the critical fixation. For this purpose, target detection probability was plotted across all five conditions as a function of target–saccade angle (i.e., the angle between target position and landing position of the subsequent saccade; in steps of 45° of arc) and saccade amplitude (ranges: $< 5°$, 5° to 10°, $> 10°$), as illustrated in Figure 3.7. Corresponding to the results of Experiment 1, target detection probability was best when saccade amplitudes were similar to the target's distance from the fovea rather independent of saccade direction. A GLMM with a logit link function and treatment contrasts tested for differences of target detection probability between the three ranges of saccade amplitude ($< 5°$, 5° to 10°, and $> 10°$ of visual angle). The range of amplitudes $< 5°$, which includes the target's distance from the fovea, served as the reference category ($b = 1.37, SE = 0.10, z = 13.72, p < 0.001$). Compared to this amplitude range, target detection probability significantly decreased for saccade amplitudes of 5° to 10° ($b = -1.41, SE = 0.11, z = -12.70, p < 0.001$) and for amplitudes $> 10°$ ($b = -1.82, SE = 0.11, z = -16.86, p < 0.001$). As in Experiment 1, the results suggest that target detection probability correlated with the amplitude rather than the direction of the planned saccade.

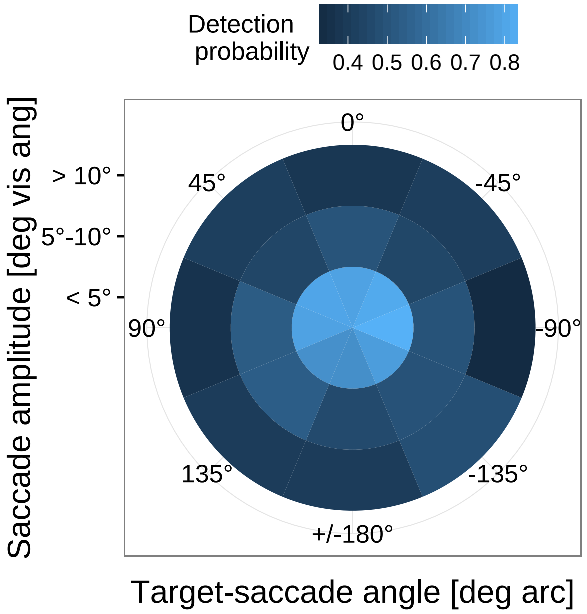

Figure 3.7: Target detection probability as a function of target–saccade angle (see Figure 3.4a for further details) and saccade amplitude in Experiment 2. Probability was computed for target–saccade angles in steps of 45° of arc and for three ranges of saccade amplitude: $< 5°$, 5° to 10°, and $> 10°$ of visual angle. The small amplitude range ($< 5°$) comprised all possible target positions at 3.5° retinal eccentricity. Lighter colors in the plot correspond to a higher probability to detect the target.

3.4 Discussion

In two experiments we investigated how the degradation of complex natural scenes in the central or the peripheral visual field affects the distribution of attention. For this purpose, participants had to detect central or peripheral targets while viewing scenes under experimental gaze-contingent degradation. Previous research suggests that the useful field of view shrinks with increasing processing demands in the central or the peripheral visual field (Crundall et al., 1999; Holmes et al., 1977; Ikeda & Takeuchi, 1975; Loschky & McConkie, 2002; Williams, 1985, 1989, 1995); consequently, performance in a peripheral detection task would be expected to decrease when either part of the scene is degraded. However, saccades are typically longer with central filtering and shorter with peripheral filtering (Foulsham et al., 2011; Laubrock et al., 2013; Loschky & McConkie, 2002; Loschky et al., 2005; Nuthmann, 2013, 2014; Nuthmann & Malcolm, 2016; Reingold & Loschky, 2002; Shioiri & Ikeda, 1989; van Diepen & Wampers, 1998; van Diepen, 2001). This saccade amplitude pattern might reflect a modified distribution of attention: a narrower attentional focus with peripheral filtering and a stronger attentional bias toward the periphery with central filtering.

To explicitly test this hypothesis, we filtered high or low spatial frequencies in the central or the peripheral visual field during scene viewing and measured the detection of peripheral (Experiment 1) or central targets (Experiment 2) that were present in the scenes during randomly chosen fixations. Assuming that saccade amplitudes reflect attentional selection, peripheral target detection was expected to decrease with peripheral but not with central filtering whereas central target detection was expected to decrease with central but not with peripheral filtering.

3.4.1 Main experimental findings

Both experiments show that peripheral spatial frequency filtering shortens saccade amplitudes and that central filtering lengthens saccade amplitudes, which is in line with previous findings (Foulsham et al., 2011; Laubrock et al., 2013; Loschky & McConkie, 2002; Loschky et al., 2005; Nuthmann, 2013, 2014; Nuthmann & Malcolm, 2016; Reingold & Loschky, 2002; Shioiri & Ikeda, 1989; van Diepen & Wampers, 1998; van Diepen, 2001). These effects were stronger when spatial frequencies that are most important for foveal analysis and peripheral target selection were missing, that is, with central low-pass and peripheral high-pass filtering. The latter finding differs from our previous finding in Laubrock et al. (2013), in which peripheral low-pass filtering shortened saccade amplitudes slightly more than peripheral high-pass filtering did. The present experiments, however, corroborate the notion that saccade amplitudes are increasingly adjusted as the potential usefulness of information decreases.

A new result reported here is a strong and specific effect of central and peripheral scene degradation on target detection in these regions. We observed that target detection probability changed according to the different filter conditions, indicating that central and peripheral scene degradation entail a modulation of attention. Peripheral filtering decreased detection probability and increased response times for peripheral targets, particularly with high-pass filtering. With central filtering, detection probability and response times for peripheral targets were similar to the unfiltered control condition with both filter types. For central targets, effects were reversed. Central filtering decreased detection probability and increased response times for central targets, particularly with high-pass filtering. With peripheral filtering, target detection was unaffected with both filter types, and response times slightly increased with high-pass filtering but not with low-pass filtering.

3.4.2 Effects of peripheral filtering: Tunnel vision

Our findings support the hypothesis that peripheral degradation of natural scenes provokes tunnel vision (Mackworth, 1965; Williams, 1988) with a reduced attentional focus that leads to a decreased perceptibility of peripheral stimuli. Thus, shorter saccade amplitudes reflect a reduced spatial extent of attention. Filtering spatial frequencies in the peripheral visual field impairs the processing of peripheral scene information and saccade target selection. With peripheral low-pass filtering, viewers took more time to process the blurred peripheral information, indicated by increased fixation durations and increased response times to peripheral targets. Furthermore, decreased saccade amplitudes together with a decreased probability to detect the peripheral target indicate a reduced attentional span with the attentional focus on the unfiltered central region. This focus became even stronger with peripheral high-pass filtering. High-pass filtering attenuates low spatial frequencies that are essential for the processing of peripheral information so that little useful information is left in the periphery. As a consequence, mean saccade amplitude shortened to about 5°, and mean fixation duration numerically decreased below the baseline, which is reasonable because the amount of potentially useful information needing to be processed was reduced. Furthermore, peripheral detection performance was markedly worse, and target response times were strongly inflated compared with low-pass filtering and the unfiltered control. Thus, shorter saccades reflected a smaller attentional span that often caused peripheral information to be ignored (hence the decrease in fixation durations with reduced time for peripheral processing) and saccade targets to be chosen from the unfiltered central visual field.

We conclude that peripheral filtering shrinks the attentional focus and thus lowers the sensitivity to peripheral stimuli in the scene. This effect strengthens as the amount of potentially useful information decreases. The detectability of target stimuli in the central visual field, however, was similar to the control condition with peripheral filtering for both filter types. Thus, peripheral filtering did not impair central target detection. However, it did not elevate detection probability above the baseline either, which is what one might expect when attention is mainly focused on the central visual field and withdrawn from the periphery. The lack of improvement in target detection is compatible with results from experiments that orthogonally varied the size of the attended area and found that, given a situation with a small stimulus and a large attention field, valid cues do not cause cueing benefits, whereas invalid cues lead to costs (Herrmann et al., 2010). Furthermore, the visibility of the target was set to 75% detection probability for each background and participant in advance, and it might not be possible for participants to exceed this threshold even with more attentional resources available for processing in the central region.

3.4.3 Effects of central filtering: Attentional bias toward the periphery

When information in the central visual field was degraded, detection probability for central targets decreased markedly and response times increased whereas detection probability for peripheral targets was unaffected. This indicates that longer saccade amplitudes with central filtering reflect a withdrawal of attention from the central visual field toward the periphery, thus decreasing the perceptibility of central stimuli. With low-pass filtering, only blurred information is left in the central visual field, impairing the identification of objects and the analysis of details. Consequently, central information is not processed exhaustively, but attention is shifted more frequently toward the unfiltered peripheral region as indicated by a reduced proportion of short saccades, an increased amount of long saccades, and a decrease in detection probability for central targets.

Surprisingly, performance in the central detection task was substantially weaker with central high-pass than with low-pass filtering. As the central visual field is very sensitive to medium and high spatial frequencies (Hilz & Cavonius, 1974; Robson & Graham, 1981), which are critical for foveal analysis, we expected processing to be easier with central high-pass filtering than with central low-pass filtering. Therefore, target detection probability was expected to be higher for high-pass filtering as attention should not be withdrawn that much from the central visual field. We suggest two post hoc explanations for the strong decrease of detection probability. First, we did not directly test for the usefulness of high or low spatial frequencies in either part of the visual field. Thus, we cannot rule out that high-pass filtering actually hampered processing more than low-pass filtering (e.g., due to the loss of contrast, luminance, and color), leading to a stronger withdrawal of attention from the central region and thus to a lower perceptibility to central targets. However, effects of central high-pass filtering on saccade amplitudes—if reflecting attentional selection—are not compatible with this interpretation as mean amplitudes were smaller than with low-pass filtering, and the amplitude distribution was similar to the one for the unfiltered control condition. We therefore suggest the following explanation. High-pass filtered information was more critical for processing than low-pass filtered information, making it more worthwhile to analyze the filtered central region. However, the processing of experimentally degraded information requires more resources than the processing of undegraded information; hence, the attentional focus during the processing of high-pass filtered central information might have been smaller than usual, concentrating strongly on foveal information processing.

An interesting new finding of the present study is that central filtering did not affect the sensitivity to peripheral targets. Thus, although foveal load presumably increased due to higher processing demands with central filtering, the size of the attentional focus was not reduced. Previous research, however, suggests that performance on peripheral detection is reduced by increased foveal load (Crundall et al., 1999; Holmes et al., 1977; Ikeda & Takeuchi, 1975; Ringer, Throneburg, Johnson, Kramer, & Loschky, 2016; Williams, 1985, 1989, 1995) because more attentional resources are needed in the central region, which shrinks the functional field of view. The present study provides no evidence for such a shrinkage with increased central processing demands. However, we induced increased foveal load in terms of a very low-level visual feature, namely the spatial frequency content of the images, which is known to be processed by cells in V1 (Shapley & Lennie, 1985; Tootell, Silverman, & De Valois, 1981). An attentional filter may thus be applied at a fairly low level of processing. Conversely, the aforementioned studies on foveal load appear to involve attentional resources at a higher, often executive, level of processing (e.g., foveal letter discrimination). Thus, the opposing results between our study and previous studies might be explained by the different levels of processing that the respective foveal tasks involved. Note, however, that the increase in fixation durations we found with central filtering suggests that the filtering effects were propagated to the cognitive level.

Our results, although suggesting that attentional resources are withdrawn from the degraded central region toward the peripheral visual field, provide no evidence for an improvement of peripheral target detection. We assume that three factors contribute to the lack of improvement in the peripheral detection task with central filtering. First, viewers were confronted with two tasks that they had to trade off against each other (scene memorization and target detection). As central filtering hampered the encoding of the central stimulus, scene memorization was more difficult, which might have interfered with peripheral target detection. Second, the peripheral visual field comprises the largest part of the scene, so any given area has only a small chance to be covered by a narrow attentional spotlight, and conversely a very diffuse spotlight of attention (i.e., a zoomed-out zoom lens of attention) would only barely raise the level across the whole field (Eriksen & Yeh, 1985; Eriksen & James, 1986; Müller, Bartelt, Donner, Villringer, & Brandt, 2003). Therefore, it is unlikely a priori that peripheral target sensitivity was elevated to a level markedly above the baseline. Third, as explained above, it is unlikely that the threshold of 75% detection probability that was determined for each background and participant in advance could be exceeded.

3.4.4 Post hoc analyses: Partial independence of saccade amplitudes and direction

As viewers were engaged in a dynamic task that allowed them to freely move their eyes around, we also investigated the coupling of attention and eye-movement planning and execution. We analyzed how the probability of covertly detecting the central or peripheral target varied as a function of size and direction of the subsequent saccade. Results for both central and peripheral targets showed a marked enhancement of target detection probability for saccade amplitudes corresponding to the target's distance from the fovea, surprisingly, with the enhancement being rather independent of saccade direction. This is compatible with a ring-shaped allocation of covert attention induced by the present task demands and might indicate that attention scans several parafoveal or peripheral locations in parallel for detecting the target, of which only one is chosen as the next saccade target. Although we obtained these results in a post hoc analysis and thus cannot disentangle cause and consequence, our findings appear to corroborate the idea that covert attentional scanning is faster than overt attention, and that it can be allocated in a ring-shaped fashion (Egly & Homa, 1984; Juola et al., 1991). Alternatively, the results might suggest that saccade amplitude and saccade direction are computed in partial independence.

3.5 Conclusions

In summary, we observed reduced saccade lengths induced by peripheral scene degradation and increased saccade lengths induced by central scene degradation. Our results lend support to the interpretation that these effects in scene-exploration behavior are created by modulations of covert attention. We have shown that increased peripheral processing demands lead to tunnel vision with a withdrawal of attention from the peripheral visual field. This effect is considerably reduced when low-frequency information is preserved in the periphery, which constitutes further evidence for the selective importance of different spatial frequency bands for saccade planning and peripheral scene analysis. Furthermore, the present findings indicate that increased central processing demands during scene viewing can lead to a withdrawal of attention from (at least part of) the central visual field and an attentional bias toward the periphery. Thus, the attentional focus does not necessarily become smaller with increasing central processing demands. Finally, we have shown in a post hoc analysis that, although saccade planning and target detection tend to go along, attention can also select locations that are not related to saccades. These results are highly relevant to theories on the coupling of attention and eye movements.

3.6 Acknowledgments

This work was funded by Deutsche Forschungsgemeinschaft (grants LA 2884/1 to J. L. and EN 471/10 to R. E.). We thank Petra Schienmann and our student assistants for their help during data collection.

Chapter 4

Spatial frequency processing in the central and peripheral visual field during scene viewing

Anke Cajar, Ralf Engbert, and Jochen Laubrock
University of Potsdam, Germany

Running head: Spatial frequency processing

published 2016 in *Vision Research, 127*, 186–197,
doi:10.1016/j.visres.2016.05.008.

Abstract

Visuospatial attention and gaze control depend on the interaction of foveal and peripheral processing. The foveal and peripheral regions of the visual field are differentially sensitive to parts of the spatial-frequency spectrum. In two experiments, we investigated how the selective attenuation of spatial frequencies in the central or the peripheral visual field affects eye-movement behavior during real-world scene viewing. Gaze-contingent low-pass or high-pass filters with varying filter levels (i.e., cutoff frequencies; Experiment 1) or filter sizes (Experiment 2) were applied. Compared to unfiltered control conditions, mean fixation durations increased most with central high-pass and peripheral low-pass filtering. Increasing filter size prolonged fixation durations with peripheral filtering, but not with central filtering. Increasing filter level prolonged fixation durations with low-pass filtering, but not with high-pass filtering. These effects indicate that fixation durations are not always longer under conditions of increased processing difficulty. Saccade amplitudes largely adapted to processing difficulty: amplitudes increased with central filtering and decreased with peripheral filtering; the effects strengthened with increasing filter size and filter level. In addition, we observed a trade-off between saccade timing and saccadic selection, since saccade amplitudes were modulated when fixation durations were unaffected by the experimental manipulations. We conclude that interactions of perception and gaze control are highly sensitive to experimental manipulations of input images as long as the residual information can still be accessed for gaze control.

4.1 Introduction

Why do we move our eyes? Due to sensory and cognitive limitations, high-acuity vision is restricted to the central 2° of the visual field, the fovea, whereas the visual periphery is rather blurry (Strasburger et al., 2011; Wertheim, 1894). As a consequence, high-velocity saccades shift the gaze about three times each second to bring regions of interest from the low-resolution periphery into the fovea for closer inspection. Two tasks are accomplished during the following fixation: fine-grained foveal information is analyzed to identify objects and details, and coarse-grained peripheral information is analyzed to select the next saccade target among competing regions of interest. Thus, visual information in the central and the peripheral visual field serve different tasks (Gilchrist, 2011).

The present study investigates how the two tasks of foveal analysis and peripheral selection are accomplished during real-world scene viewing when fine-grained or coarse-grained information is selectively attenuated in the central or the peripheral visual field. Inherently this also sheds light on the question to what degree central and peripheral vision contribute to spatial and temporal aspects of eye-movement behavior. The issue can be tackled by attenuating high or low spatial frequencies in the central or the peripheral visual field. High spatial frequencies provide the fine-grained information of an image and low spatial frequencies provide the coarse-grained information of an image. High-pass filters preserve high spatial frequencies and attenuate low spatial frequencies; with low-pass filters, it is vice versa. Information can be selectively altered in either the central or the peripheral part of the visual field by applying a gaze-contingent window of arbitrary size that moves with the current gaze position of the viewer in real-time during scene inspection (McConkie & Rayner, 1975; Rayner & Bertera, 1979). Spatial frequencies are filtered either inside or outside the gaze-contingent window with central or peripheral filtering respectively, while the other region of the scene remains unchanged.

Previous research on this topic is rather scant and has mostly been focused on the effects of peripheral low-pass filtering. Corresponding studies indicate that spatial-frequency filtering impairs scene processing, as viewers' performances in several tasks decrease with filtering. For example, when searching for objects in scenes, search accuracy decreases and search times increase with peripheral as well as with central low-pass filtering; these effects get stronger as filter level and filter size increase (Loschky & McConkie, 2002; Nuthmann, 2014). Furthermore, the probability to detect target stimuli in low-pass or high-pass filtered scene regions decreases and response times for detected targets increase (Cajar, Schneeweiß, Engbert, & Laubrock, 2016). Central low-pass filtering has also been shown to decrease response accuracy to memory questions about scenes (Cajar, Schneeweiß, et al., 2016). These findings suggest that the processing difficulty of a scene increases with spatial-frequency filtering, and increases more the larger or stronger the filter gets.

In agreement with the decrease in task performance, eye-movement behavior has been reported to deviate progressively from viewing behavior in unfiltered scenes as spatial-frequency filtering increases processing difficulty. Studies consistently show that viewers prefer unfiltered scene regions as saccade targets. Peripheral filtering shortens mean saccade amplitudes, since viewers tend to keep their gaze inside the unfiltered central region and avoid longer saccades to the filtered periphery (Cajar, Schneeweiß, et al., 2016; Foulsham et al., 2011; Laubrock et al., 2013; Loschky & McConkie, 2002; Loschky et al., 2005; Nuthmann, 2013, 2014; Shioiri & Ikeda, 1989). Central filtering, on the other hand, lengthens mean saccade amplitudes, since viewers tend to place fewer saccades inside the filtered center and make more long saccades to the periphery (Cajar, Schneeweiß, et al., 2016; Laubrock et al., 2013; Nuthmann, 2014). With both central and peripheral low-pass filtering, the effects get larger with increasing filter size (Loschky & McConkie, 2002; Nuthmann, 2013, 2014) and filter level (Loschky & McConkie, 2002). Thus, saccadic selection is modulated more and more as processing difficulty increases. It has been shown recently that these changes in saccade amplitudes go along with corresponding changes in visuospatial attention (Cajar, Schneeweiß, et al., 2016).

Fixation duration also varies with visual-cognitive processing and usually increases as the acquisition of information from the scene becomes more difficult (Henderson, 2003; Nuthmann et al., 2010). Thus, studies show that fixation durations increase with spatial-frequency filtering of the entire scene (Glaholt et al., 2013; Henderson et al., 2014; Mannan et al., 1995) as well as with central low-pass filtering (Cajar, Schneeweiß, et al., 2016; Nuthmann, 2014) and with peripheral low-pass filtering (Cajar, Schneeweiß, et al., 2016; Laubrock et al., 2013; Loschky & McConkie, 2002; Loschky et al., 2005; Nuthmann, 2013, 2014; Parkhurst et al., 2000; van Diepen & Wampers, 1998). Fixations also increasingly prolong with increasing low-pass filter size (Nuthmann, 2013, 2014; Parkhurst et al., 2000). However, Loschky and colleagues found that increasing filter level with detectable peripheral low-pass filtering hardly affected fixation durations (Loschky & McConkie, 2002; Loschky et al., 2005). In summary, previous research suggests that eye-movement behavior is increasingly modulated as visual-cognitive processing difficulty increases due to spatial-frequency filtering.

In contrast, we recently found evidence in two studies (Cajar, Schneeweiß, et al., 2016; Laubrock et al., 2013) that fixation durations are not always longer under conditions of increased processing difficulty. In both studies, high-pass filters or low-pass filters were applied to either the central or the peripheral part of the visual field during the viewing of color (Laubrock et al., 2013) or grayscale (Cajar, Schneeweiß, et al., 2016) real-world scenes. We assumed that scene processing would be most difficult with central low-pass and peripheral high-pass filtering, as these conditions strongly attenuate the critical spatial frequencies for foveal analysis (high spatial frequencies) and peripheral target selection (low spatial frequencies) respectively. In both studies, however, mean fixation durations increased most with central high-pass and peripheral low-pass filtering, which were expected to be less disruptive for processing. Central low-pass and peripheral high-pass filtering involved shorter mean fixation durations, often similar to the mean fixation duration in the unfiltered control condition. The results suggest that viewers invested more processing time when the information left after filtering was useful enough to accomplish the task at hand (foveal analysis, peripheral selection) in a reasonable amount of time; when visual-cognitive processing became too difficult to make an investment of more processing time worthwhile default timing, that is, stimulus-independent random timing of saccades was adapted. To account for these effects, we developed a computational model in which fixation durations are controlled by the dynamical interaction of foveal and peripheral processing (Laubrock et al., 2013). The model assumes that foveal and peripheral information processing evolve in parallel and independently from one another during fixation, a notion that was recently corroborated by an experimental study (Ludwig, Davies, & Eckstein, 2014).

4.1.1 The present study

Most studies on the effects of gaze-contingent spatial-frequency filtering on eye movements during scene viewing applied peripheral low-pass filters. There is only little research on the effects of central filtering (Nuthmann, 2014) and high-pass filtering (van Diepen & Wampers, 1998) on eye-movement behavior. Our own studies (Cajar, Schneeweiß, et al., 2016; Laubrock et al., 2013) were the first to investigate the effects of central and peripheral high-pass and low-pass filtering within the same experiment. Moreover, to our knowledge, there is no study on the effects of varying filter level and filter size with high-pass filtering. The investigation of the latter is interesting on its own. In addition, the aforementioned effects on fixation durations in our previous studies with varying filter type (low-pass/high-pass) in the central or peripheral visual field are striking—they raise the question how fixation durations adapt to processing difficulty due to spatial-frequency filters of varying filter level and filter size. The present study investigated this question in two experiments.

In both experiments, participants inspected real-world scenes in preparation for a memory task while high or low spatial frequencies were filtered either in the central or the peripheral visual field. Gaze-contingent filtering was compared with control conditions that presented scenes either unfiltered or entirely low-pass or high-pass filtered. In Experiment 1, the level of filtering (i.e., the cutoff frequency) varied between trials using weak, moderate, or strong high-pass or low-pass filters. Processing difficulty was assumed to increase from weak to strong filters. In Experiment 2, filter level was constant, but the size of the filter (i.e., the size of the gaze-contingent window) varied—the filter either subtended a small, medium, or large region of the central or the peripheral visual field. Processing difficulty was assumed to increase from small to large filters. The experiments tested for the effects of filtering on task performance, fixation durations, and saccade amplitudes.

For both experiments, we expected saccade amplitudes to increasingly deviate from normal viewing behavior with increasing processing difficulty. Compared with unfiltered scene viewing, amplitudes were expected to increase with central filtering and decrease with peripheral filtering, particularly when critical spatial frequencies were attenuated (i.e, with central low-pass filtering and peripheral high-pass filtering). These effects were expected to grow with increasing filter level (Experiment 1) or filter size (Experiment 2). For fixation durations, we expected an increase as long as potentially more useful spatial frequencies were preserved and information uptake was relatively easier, that is, with central high-pass and peripheral low-pass filtering. Furthermore, durations were expected to increase as filters became stronger or larger; however, as observed previously for varying filter type (Cajar, Schneeweiß, et al., 2016; Laubrock et al., 2013) default timing might be adapted with strong or large filters that attenuate most useful spatial frequencies and strongly impede information uptake.

4.2 Method

4.2.1 Participants

Participants were students from the University of Potsdam. For Experiment 1, thirty-two people were tested (10 male, mean age: 21.9 years); for Experiment 2, another thirty-two people, who did not participate in Experiment 1, were tested (11 male, mean age: 22.1 years). They received course credit or fifteen euro for participation in the experiment. All participants had normal or corrected-to-normal vision and normal color discrimination. Participants gave their written informed consent prior to the experiments, which were carried out in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki).

4.2.2 Apparatus

Stimuli were presented on a 20″ Mitsubishi DiamondPro 2070 CRT monitor with a resolution of 1024 × 768 pixels and a refresh rate of 120 Hz. A head-chin rest was used to reduce participants' head movements and ensure a constant viewing distance of 60 cm. Gaze position of the right eye was recorded during binocular viewing using an EyeLink 1000 tower mount system (SR Research, Ontario, Canada) with a sampling rate of 1000 Hz. Stimulus presentation was controlled with MATLAB® (version 2009b; The Mathworks, Natick, MA) using the OpenGL-based Psychophysics Toolbox (PTB3; Brainard, 1997; Kleiner et al., 2007; Pelli, 1997) and the Eyelink Toolbox (Cornelissen et al., 2002). The highest and lowest possible spatial frequency the monitor could display was 13.2 c/° and 0.03 c/° respectively. System latency from eye movement to screen update (as a sum of frame duration, tracker latency, and eye-velocity computation) was smaller than or equal to 11 ms.

4.2.3 Stimuli

Each experiment consisted of two sessions presenting 240 different color photos of outdoor real-world scenes (120 scenes per session). The same scenes were used in Experiments 1 and 2. Scenes were displayed at a resolution of 1024 × 768 pixels and a size of 38.7° × 29.0°. They were low-pass filtered in one session and high-pass filtered in the other session. The filtered version of each scene was prepared in advance. Filtering with first-order Butterworth filters was realized in the frequency domain after a Fourier transform of the stimulus, and the filtered image was transformed back into the spatial domain.

For gaze-contingent filtering in the central or the peripheral visual field, a foreground and a background image were merged in real-time using alpha blending. With peripheral high-pass filtering, for example, the foreground image was the original scene and the background image was the high-pass filtered version of the scene. A 2D hyperbolic tangent with a slope of 0.06 was used as a blending function for creating the alpha mask. The inflection point of the function corresponded to the radius of the gaze-contingent window. The alpha mask was centered at the current gaze position and gave the transparency (i.e., the weighting) of the central foreground image at each point. At the point of fixation, only the foreground image was visible; with increasing eccentricity, the peripheral background image was weighted more strongly until it was fully visible (cf. Cajar, Schneeweiß, et al., 2016). To avoid jitter due to fixational eye movements, the gaze-contingent display was updated only when the instantaneous eye velocity exceeded 186°/s or after a duration of 200 ms.

4.2.3.1 Specifics for stimuli in Experiment 1

In Experiment 1, the gaze-contingent moving window had a constant radius of 3.75°. For the three filter levels, filters with cutoff frequencies of 1.26 c/°, 3.16 c/°, and 7.94 c/° were used; spatial frequencies were attenuated above these thresholds for low-pass filtering and below these thresholds for high-pass filtering. Weak filters preserved a relatively wide band of spatial frequencies and strong filters preserved a relatively narrow band of spatial frequencies; moderate filters were in between. For low-pass filtering, the weak filter had a cutoff frequency of 7.94 c/° and the strong filter had a cutoff frequency of 1.26 c/°; for high-pass filtering, the assignment was reversed. All filters were above detection threshold (cf. Loschky et al., 2005). Example stimuli for Experiment 1 are illustrated in Figure 4.1a.

To give the reader a rough estimate how far off into the periphery the cutoff frequencies were visible, we calculated the maximal eccentricities they were visible at according to the cortical magnification principle. Contrast sensitivity for gratings with spatial frequencies measured in cycles/mm of cortex, that is, c/° divided by the cortical magnification factor M, is independent of eccentricity (Rovamo, Virsu, & Näsänen, 1978; Rovamo & Virsu, 1979; Virsu, 1979). Rovamo and Virsu (1979, p. 498) provide four formulas to calculate M for different eccentricities—one for each half-meridian of the visual field (superior, inferior, nasal, and temporal), since cortical magnification differs between them. Using these formulas, we estimated the eccentricities at which the cutoff frequencies used in the present experiment became invisible on the superior and temporal half-meridians (along which contrast sensitivity is worst and best respectively). Assuming that gratings with a spatial frequency greater than 3.5 c/mm of cortex are invisible (Rovamo & Virsu, 1979, Fig. 4), the cutoff frequencies should become invisible at the following eccentricities: 36.8° and 63.3° for spatial frequencies of 1.26 c/°; 18.5° and 28.5° for frequencies of 3.16 c/°; 7.9° and 11.6° for frequencies of 7.94 c/°.

4.2.3.2 Specifics for stimuli in Experiment 2

In Experiment 2, filter level was constant with a cutoff frequency of 1.0 c/° for low-pass filtering and 10.0 c/° for high-pass filtering, spatial frequencies at which the parvo and magno cells of the lateral geniculate nucleus are near their peak contrast sensitivity (Derrington & Lennie, 1984). According to the aforementioned derivation, these cutoff frequencies approximately become invisible on the superior and temporal half-meridians at the following eccentricities: 42.4° and 75.3° for 1.0 c/°; 6.3° and 9.3° for 10.0 c/°. The size of the filtered region was varied by varying the radius of the gaze-contingent window, with values of 1.75°, 3.75°, and 5.75° for the small, medium, and large gaze-contingent window respectively. These window sizes roughly extended to the foveal, the parafoveal, and the peripheral visual field (see Larson & Loschky, 2009). Thus, small central filters mainly degraded foveal vision whereas large central filters degraded foveal and parafoveal vision; small peripheral filters, on the other hand, degraded only peripheral vision whereas large peripheral filters degraded peripheral and parafoveal vision. Increasing window size from small to large therefore increased the size of the filtered region with central filtering, but decreased the size of the filtered region with peripheral filtering. Example stimuli for Experiment 2 are illustrated in Figure 4.1b.

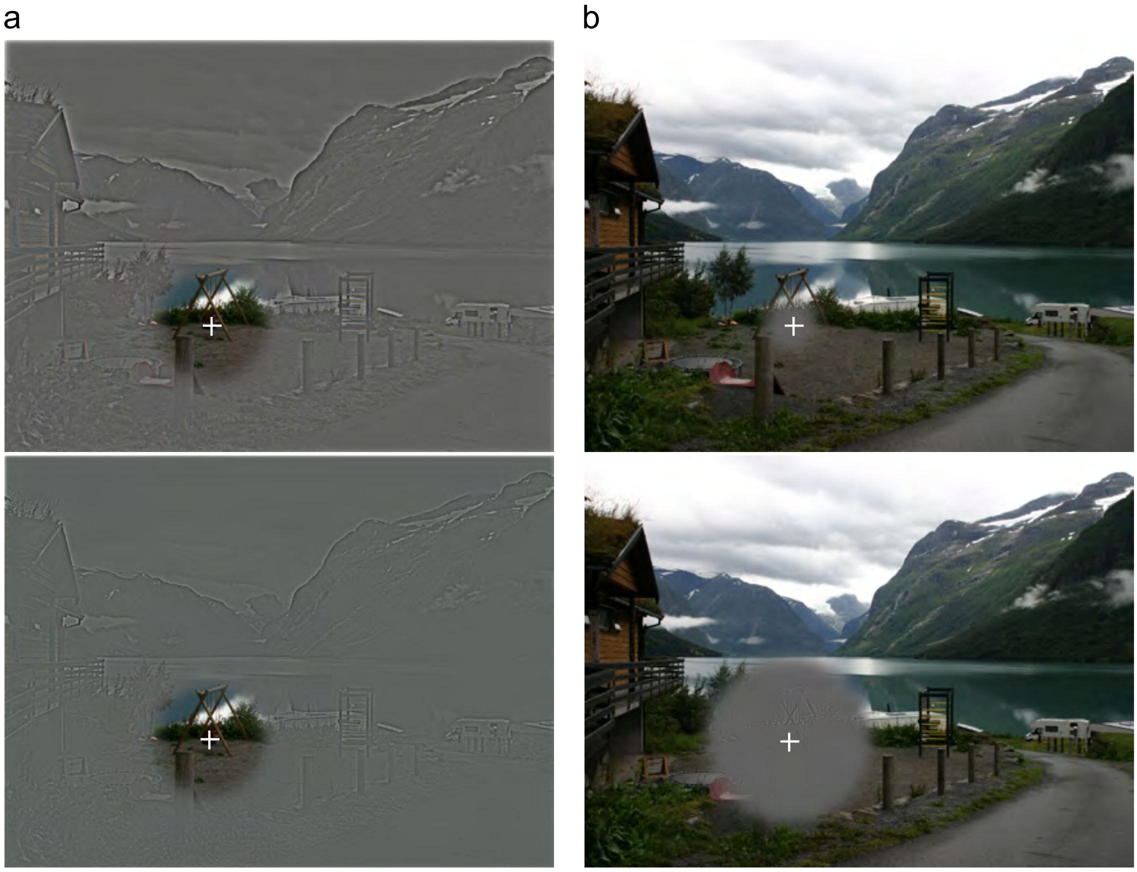

Figure 4.1: Illustration of four experimental conditions. The white cross illustrates the current gaze position of the viewer. (a) Two filter conditions of Experiment 1: Weak peripheral high-pass filter (top) and moderate peripheral high-pass filter (bottom). (b) Two filter conditions of Experiment 2: Small central high-pass filter (top) and large central high-pass filter (bottom). Note that the illustrated scenes are at a smaller scale than presented in the experiments and therefore do not faithfully reproduce the actual filters used in the experiments.

4.2.4 Design

Each experiment consisted of two sessions, one for each filter type (low-pass, high-pass). As each participant completed both sessions, session order was counterbalanced so that half of the participants started with the low-pass session and the other half with the high-pass session. In each session, two filter locations (central visual field, peripheral visual field) were crossed with either three filter levels (weak, moderate, strong) in Experiment 1 or with three filter sizes (small, medium, large) in Experiment 2, yielding six experimental conditions. Two control conditions were added as a lower and upper baseline, presenting scenes either without any filtering (unfiltered control) or entirely filtered (filtered control). For the filtered control condition in Experiment 1, the moderate filter was used. A Latin square design assured counterbalancing of condition–image assignments across participants. Images were presented in random order.

4.2.5 Procedure

For each experiment, data were collected in two one-hour sessions. The eye-tracker was calibrated at the beginning of a session and after every fifteen trials. Each trial started with a fixation point in the center of the screen. The scene was revealed after the point had been fixated for at least 150 ms within two seconds from trial start; otherwise a recalibration was automatically scheduled.

Each session started with four practice trials to acquaint participants with the gaze-contingent display. Each scene was presented for twelve seconds in one of the eight conditions. Viewers were asked to explore each scene carefully in preparation for a memory question on scene content with three response alternatives presented after the trial. The questions were constructed to be rather difficult to answer, thus encouraging participants to inspect the scenes carefully throughout the entire twelve seconds of scene presentation. There were no questions regarding colors in a scene. Questions typically asked about the presence or absence of certain objects in the scene (e.g., “Which of the three following objects was not present in the scene?”), about the location of objects (e.g., “Where was the suitcase standing?”), or about the number of certain objects (e.g., “How many people were present in the scene?”). The memory question for the scene in Figure 4.1, for example, was “Which object was seen on the playground?”, with the response alternatives “Sandbox”, “Trampoline”, or “Slide” (correct answer: “Trampoline”).

4.2.6 Data preparation

Saccades were detected using a velocity-based adaptive algorithm (Engbert & Kliegl, 2003; Engbert & Mergenthaler, 2006). A total of 480 trials (6%) of Experiment 1 and 478 trials (6%) of Experiment 2 were removed owing to poor recording or too much data loss. Single fixations and saccades were removed if they neighbored eye blinks or were outside of the monitor area. The first and the last event of a trial were also excluded from analyses, since they were associated with scene onset and offset. Glissades following a saccade were assigned to the saccade; if more than one glissade followed a saccade, the glissades and their adjacent fixation and saccade were removed (cf. Cajar, Schneeweiß, et al., 2016; Laubrock et al., 2013). In total, 273,716 fixations and 283,692 saccades were left for analyses of Experiment 1; 256,711 fixations and 266,912 saccades were left for analyses of Experiment 2.

4.2.7 Data analyses

For each experiment, data from both sessions were merged for the analyses. Fixation durations and saccade amplitudes were analyzed using linear mixed-effects models (LMM), and task performance was analyzed using binomial generalized linear mixed-effects models (GLMM) with a logit link function as implemented in the *lme4* package (Bates, Maechler, et al., 2015); this package is supplied in the *R* system for statistical computing (version 3.2.3; R Core Team, 2015). In addition to fixed effects for the experimental manipulations, (G)LMMs also account for random effects (i.e., variance components) due to differences between participants and scenes, which reduces unexplained variance. Fixed-effects parameters were estimated via user-defined contrasts testing for eight main effects and seven interaction effects for the three experimental factors (filter type, filter location, filter level/size) and the two control conditions.

All GLMM analyses provide regression coefficients, standard errors, z-values, and p-values for fixed effects. LMM analyses only provide regression coefficients, standard errors, and t-values, because the degrees of freedom are not known exactly for LMMs. For large data sets as in the present experiments, however, the t-distribution has converged to the standard normal distribution for all practical purposes (Baayen et al., 2008, Note 1). Therefore, t-statistics exceeding an absolute value of 1.96 were considered statistically significant on the two-tailed 5% level.

Because distributions of fixation durations and saccade amplitudes were positively skewed, both variables were transformed before model fitting to approximate normally distributed model residuals. To find a suitable transformation, the optimal λ-coefficient for the Box-Cox power transformation (Box & Cox, 1964) was estimated with the *boxcox* function of the *MASS* package (Venables & Ripley, 2002), with $y(\lambda) = (y^\lambda - 1)/\lambda$, if $\lambda \neq 0$ and $log(y)$, if $\lambda = 0$ (cf. Cajar, Schneeweiß, et al., 2016). For fixation durations, λ was near zero for both experiments (λ= 0.02 and λ= 0.06 for Experiments 1 and 2 respectively), so the log-transformation was chosen; for saccade amplitudes, $\lambda = 0.22$ (Experiment 1) and $\lambda = 0.18$ (Experiment 2) was used.

Raw data and analyses for both experiments are available at the Potsdam Mind Research Repository (PMR2) at `http://read.psych.uni-potsdam.de/index.php?option=com_content&view=article&id=149`.

4.3 Results of Experiment 1

4.3.1 Fixation durations

Mean fixation durations for all conditions are illustrated in Figure 4.2a. Consistent with the notion that fixation durations adapt to visual-cognitive processing difficulty, mean fixation durations increased when the scene was partly or entirely filtered, $b = 4.37 \cdot 10^{-2}, SE = 2.36 \cdot 10^{-3}, t = 18.5$. Interestingly, on average, filtering only the central or the peripheral part of the visual field increased mean fixation durations more than filtering of the entire scene did ($b = 1.12 \cdot 10^{-2}, SE = 2.44 \cdot 10^{-3}, t = 4.6$), an effect mainly driven by the strong increase of fixation durations with moderate and strong peripheral low-pass filters compared with the filtered controls (see Figure 4.2a). Comparing the filtered controls showed longer fixation durations with low-pass than with high-pass filtered scenes ($b = 2.35 \cdot 10^{-2}, SE = 5.51 \cdot 10^{-3}, t = 4.3$).

Regarding gaze-contingent central and peripheral filtering, significant main effects of filter type ($b = 2.29 \cdot 10^{-2}, SE = 3.64 \cdot 10^{-3}, t = 6.3$) and filter location ($b = 1.56 \cdot 10^{-2}, SE = 1.84 \cdot 10^{-3}, t = 8.5$) indicated longer fixation durations with low-pass than with high-pass filtering and with peripheral than with central filtering. These main effects, however, were qualified by a strong interaction of filter type and filter location, $b = 2.79 \cdot 10^{-2}, SE = 1.84 \cdot 10^{-3}, t = 15.2$. First, filter type had almost no effect with central filtering, but a strong effect with peripheral filtering where fixation durations increased markedly with low-pass, but not with high-pass filtering. Second, central filtering provoked longer fixations than peripheral high-pass filtering, but shorter fixations than peripheral low-pass filtering.

Main effects of filter level indicated longer fixation durations with stronger filters—fixations were longer with moderate and strong filters than with weak filters ($b = 2.20 \cdot 10^{-2}, SE = 1.94 \cdot 10^{-3}, t = 11.3$) and longer with strong filters than with moderate filters ($b = 9.01 \cdot 10^{-3}, SE = 2.26 \cdot 10^{-3}, t = 4.0$). However, these effects of filter level were only pronounced with low-pass filtering, as indicated by interactions of filter level and filter type ($b = 1.83 \cdot 10^{-2}, SE = 1.94 \cdot 10^{-3}, t = 9.4$ for weak versus moderate and strong filters, and $b = 8.37 \cdot 10^{-3}, SE = 2.26 \cdot 10^{-3}, t = 3.7$ for moderate versus strong filters). Interestingly, filter level hardly affected fixation durations with high-pass filtering, regardless of filter location. Filter level had the biggest effect with peripheral low-pass filtering, where mean fixation durations increased markedly from weak to moderate and strong filters. Therefore, an interaction of filter location and filter level indicated a stronger increase of fixation durations from weak to moderate and strong filters with peripheral filtering than with central filtering, $b = 8.75 \cdot 10^{-3}, SE = 1.94 \cdot 10^{-3}, t = 4.5$. All variance components and fixed effects of the LMM can be found in the Appendix (Table 4.1).

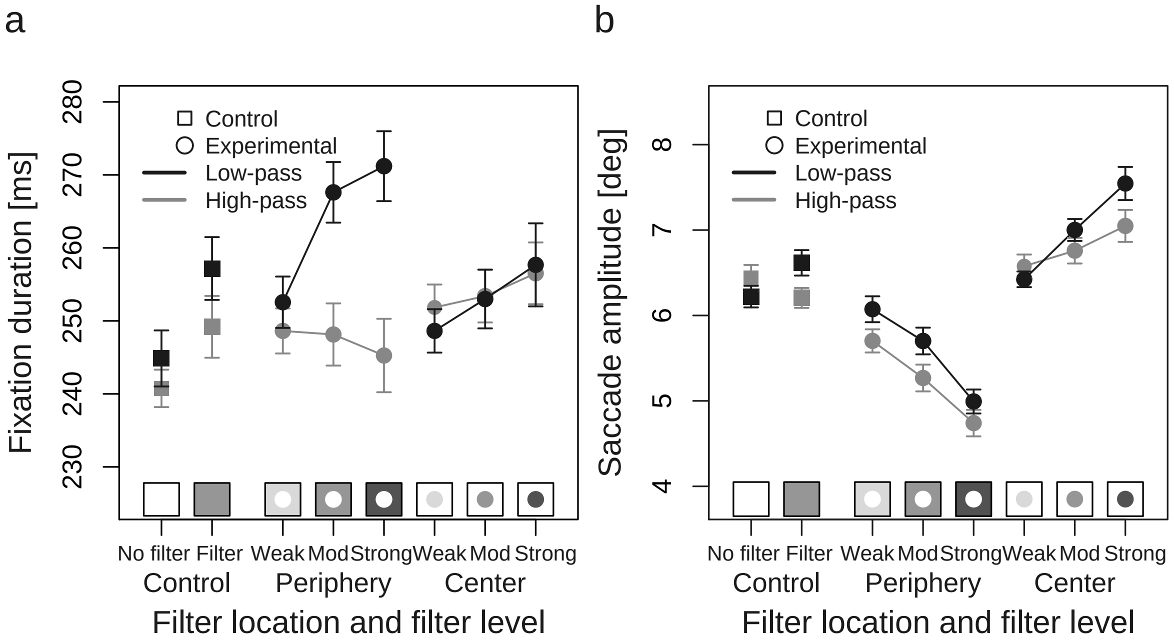

Figure 4.2: Eye-movement behavior in Experiment 1. (a) Mean fixation durations. (b) Mean saccade amplitudes. Error bars are 95% within-subject confidence intervals. The boxes at the bottom of the plots illustrate the different filter conditions—the gray area indicates the filtered part and the white area indicates the unfiltered part of the visual field; the gray level from light to dark indicates the level of filtering from weak to strong.

4.3.2 Saccade amplitudes

Mean saccade amplitudes for all conditions are illustrated in Figure 4.2b. Effects turned out as expected. There was a strong main effect of filter location ($b = 8.14 \cdot 10^{-2}, SE = 1.11 \cdot 10^{-3}, t = 73.54$), as mean saccade amplitudes were longer with central than with peripheral filtering. More precisely, central filtering elicited longer saccades than the control conditions, whereas peripheral filtering elicited shorter saccades than the control conditions. This indicates a preference for unfiltered scene regions as saccade targets. Due to this opposing viewing behavior with central and peripheral filtering, there was no main effect of filter level, but strong interactions of filter level and filter location ($b = -3.67 \cdot 10^{-2}, SE = 1.17 \cdot 10^{-3}, t = -31.33$ for weak versus moderate and strong filters, and $b = -2.65 \cdot 10^{-2}, SE = 1.36 \cdot 10^{-3}, t = -19.48$ for moderate versus strong filters)—as the filter became stronger, amplitudes increasingly lengthened with central filtering, but increasingly shortened with peripheral filtering. Thus, saccade target selection deviated more and more from normal as processing difficulty increased. Accordingly, low-pass filtering led to longer saccades than high-pass filtering with both filter locations ($b = 1.31 \cdot 10^{-2}, SE = 3.28 \cdot 10^{-3}, t = 3.98$)—because low spatial frequencies are more useful than high frequencies for peripheral target selection, more long saccades occurred with peripheral low-pass filtering; however, low spatial frequencies are less useful than high frequencies for detailed foveal analysis, so that fewer short saccades occurred with central low-pass filtering. The difference between low-pass and high-pass filters was more pronounced with peripheral than with central filtering, leading to an interaction of filter type and filter location ($b = 1.01 \cdot 10^{-2}, SE = 1.11 \cdot 10^{-3}, t = 9.13$). With peripheral filtering, filter level had a similar effect on saccade amplitudes with low-pass and high-pass filtering; with central filtering, on the other hand, filter level had a stronger effect with low-pass than with high-pass filtering. This is reflected by three-way interactions between filter type, filter location, and filter level, with $b = -1.01 \cdot 10^{-2}, SE = 1.17 \cdot 10^{-3}, t = -8.62$ for weak versus moderate and strong filters, and $b = -7.41 \cdot 10^{-3}, SE = 1.36 \cdot 10^{-3}, t = -5.45$ for moderate versus strong filters. Saccade amplitudes were least affected by weak low-pass or high-pass filters, showing that these filters preserved a sufficient amount of high or low spatial frequencies respectively for near-normal spatial viewing behavior.

Saccade amplitudes in the filtered control conditions were similar to amplitudes in the unfiltered control condition, with low-pass filtered scenes provoking slightly longer saccades than their unfiltered control and high-pass filtered scenes provoking slightly shorter saccades than their unfiltered control (see Figure 4.2b). Compatible with the results for central and peripheral filtering, completely low-pass filtered scenes led to longer saccades than high-pass filtered scenes, $b = 2.36 \cdot 10^{-2}, SE = 4.12 \cdot 10^{-3}, t = 5.73$. All variance components and fixed effects of the LMM can be found in the Appendix (Table 4.2).

4.3.3 Task performance

Mean proportion of correct answers to the memory questions are illustrated in Figure 4.3a. Task performance decreased when the scene was partly or entirely filtered, $b = -2.55 \cdot 10^{-1}, SE = 9.37 \cdot 10^{-2}, z = -2.72, p = .006$. Performance decreased with moderate and strong filters compared to weak filters, $b = -2.11 \cdot 10^{-1}, SE = 7.42 \cdot 10^{-2}, z = -2.84, p = .004$. An interaction between filter type and moderate versus strong filters ($b = 1.74 \cdot 10^{-1}, SE = 8.36 \cdot 10^{-2}, z = 2.08, p = .038$) occurred because strong filters interfered more with task performance than moderate filters with low-pass filtering, but not with high-pass filtering. There were no other significant effects.

4.3.4 Summary

In most conditions eye-movement behavior was increasingly modulated by increasing processing difficulty due to spatial-frequency filtering. Task performance decreased when scenes were filtered and decreased more as filters became stronger. Fixation durations increased with filtering. They were markedly longer with peripheral low-pass filtering than with peripheral high-pass filtering, suggesting that more processing time was invested when the degraded information could still be used efficiently for gaze control. Increasing filter level prolonged fixation durations with low-pass filtering, but had almost no effect with high-pass filtering, which we interpret as default timing. Instead, viewing behavior was adapted to stronger high-pass filters by modulation of saccade amplitudes.

As expected, saccades lengthened with central filtering and shortened with peripheral filtering compared to the unfiltered control condition, indicating a preference for unfiltered scene regions as saccade targets. Effects scaled with filter type and filter level. First, saccade amplitudes deviated more from normal viewing behavior when the available spatial frequencies were less useful for saccade target selection or foveal analysis (i.e., with peripheral high-pass filtering and central low-pass filtering respectively). Second, amplitudes monotonically lengthened with central filtering and shortened with peripheral filtering as filter level increased. Thus, the stronger the filter, the greater the bias to avoid the filtered region.

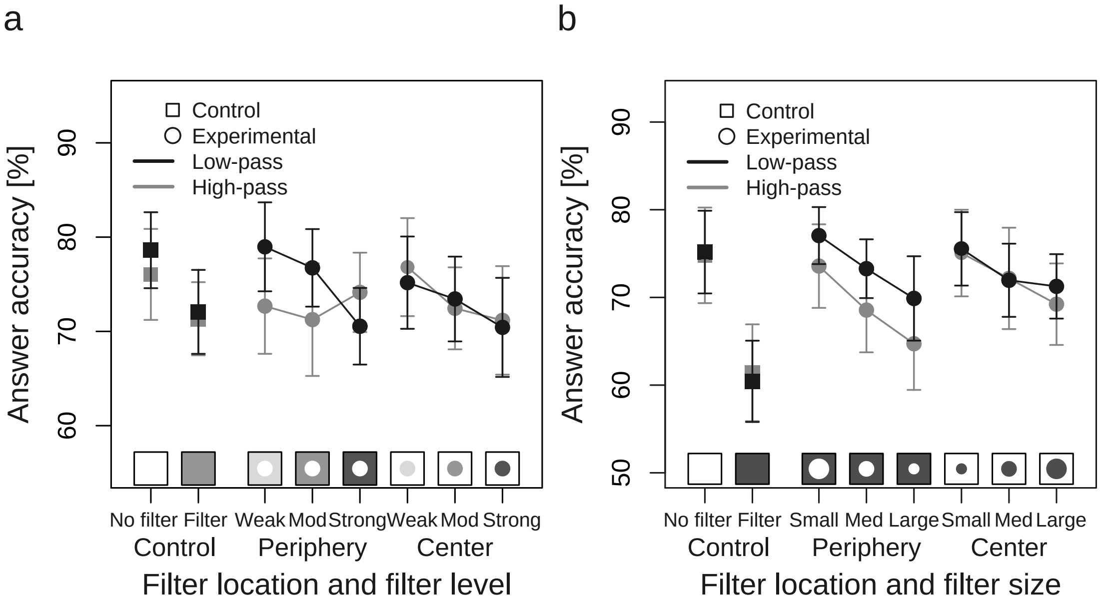

Figure 4.3: Mean proportions of correct answers to the memory questions in Experiment 1 (a) and Experiment 2 (b). Error bars are 95% within-subject confidence intervals. The boxes at the bottom of the plots illustrate the different filter conditions—the gray area indicates the filtered part and the white area indicates the unfiltered part of the visual field. The gray level from light to dark in (a) indicates the level of filtering from weak to strong.

4.4 Results of Experiment 2

4.4.1 Fixation durations

Mean fixation durations for all conditions are illustrated in Figure 4.4a. Fixation durations increased when the scene was partly or entirely filtered compared with the unfiltered control conditions, $b = 5.41 \cdot 10^{-2}, SE = 2.49 \cdot 10^{-3}, t = 21.76$. Comparing the filtered controls revealed longer fixation durations with low-pass filtering than with high-pass filtering, $b = 1.91 \cdot 10^{-2}, SE = 5.78 \cdot 10^{-3}, t = 3.31$. In contrast to Experiment 1, completely filtered scenes provoked longer mean fixation durations than central or peripheral filtering did ($b = 3.21 \cdot 10^{-2}, SE = 2.66 \cdot 10^{-3}, t = 12.09$), but fixation durations were again longer with peripheral than with complete low-pass filtering when the filter was large (see Figure 4.4a).

With central and peripheral spatial-frequency filtering the LMM revealed similar effects as in Experiment 1. Main effects of filter type ($b = 2.99 \cdot 10^{-2}, SE = 3.54 \cdot 10^{-3}, t = 8.44$) and filter location ($b = 1.68 \cdot 10^{-2}, SE = 1.92 \cdot 10^{-3}, t = 8.74$) emerged, indicating longer mean fixation durations with low-pass filtering than with high-pass filtering and longer mean fixation durations with peripheral filtering than with central filtering. These main effects were qualified by a strong interaction of filter type and filter location ($b = 5.37 \cdot 10^{-2}, SE = 1.92 \cdot 10^{-3}, t = 27.96$), as fixation durations were longer with central high-pass and peripheral low-pass filtering than with central low-pass and peripheral high-pass filtering. Consistent with Experiment 1, the difference between low-pass and high-pass filtering was larger with peripheral than with central filtering.

Most remarkable was the effect of filter size. Two main effects indicated that fixation durations increased as filters became larger ($b = 3.77 \cdot 10^{-2}, SE = 2.02 \cdot 10^{-3}, t = 18.64$ for small versus medium and large filters, and $b = 4.69 \cdot 10^{-2}, SE = 2.37 \cdot 10^{-3}, t = 19.77$ for medium versus large filters). Interestingly, this increase occurred only with peripheral filtering, particularly from medium to large filters—with central filtering, filter size had no effect on fixation durations. This is reflected in two strong interactions of filter location and filter size ($b = 4.66 \cdot 10^{-2}, SE = 2.02 \cdot 10^{-3}, t = 23.05$ for small versus medium and large filters, and $b = 4.42 \cdot 10^{-2}, SE = 2.37 \cdot 10^{-3}, t = 18.66$ for medium and large filters). Thus, although fixation durations increased with spatial-frequency filtering in the foveal region (small central filter) compared with the unfiltered control, a further extension of the filter to the parafovea did not increase fixation durations any further.

Figure 4.4a shows that, similar to Experiment 1, fixation durations were closest to the unfiltered control condition with central low-pass filtering, peripheral high-pass filtering, and peripheral low-pass filtering with the smallest filter. With small and medium peripheral high-pass filters mean fixation durations even dropped below the mean of the unfiltered control condition. All variance components and fixed effects of the LMM can be found in the Appendix (Table 4.3).

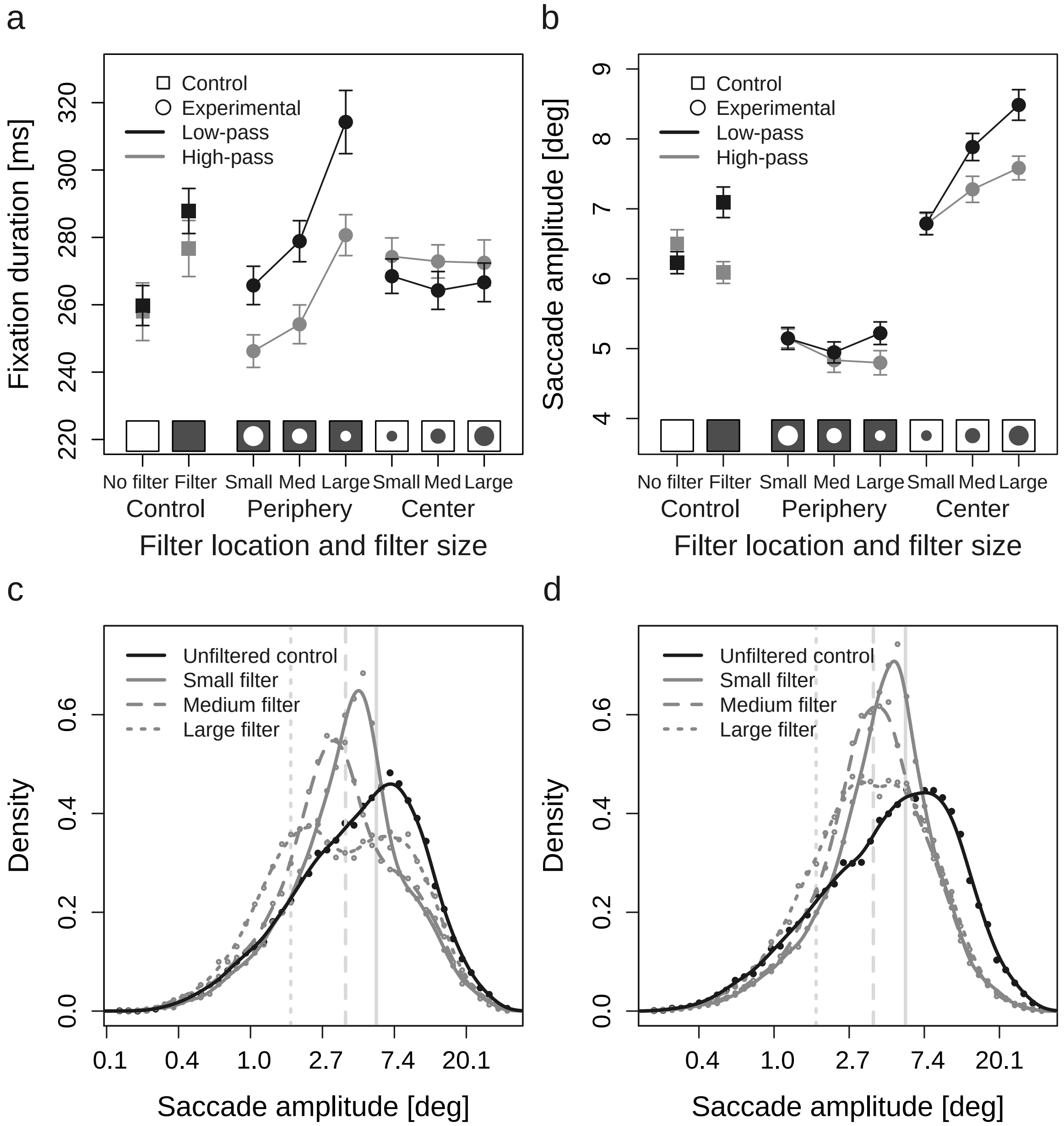

Figure 4.4: Eye-movement behavior in Experiment 2. (a-b) Mean fixation durations (a) and mean saccade amplitudes (b). Error bars are 95% within-subject confidence intervals. The boxes at the bottom of the plots illustrate the different filter conditions—the dark area indicates the filtered part and the white area indicates the unfiltered part of the visual field. (c-d) Distributions of saccade amplitudes for peripheral low-pass filtering (c) and peripheral high-pass filtering (d). Note that the abscissa is log-scaled. Lines represent kernel density estimates using a Gaussian kernel, as implemented in the R function *density*. The area under each curve adds up to one. Vertical gray lines represent the gaze-contingent window radii of 1.75° (dotted line), 3.75° (dashed line), and 5.75° (solid line) corresponding to large, medium, and small peripheral filters respectively.

4.4.2 Saccade amplitudes

Mean saccade amplitudes for all conditions are illustrated in Figure 4.4b. Results largely turned out as in Experiment 1 and showed adaptation of saccade amplitudes to increasing processing difficulty. A strong main effect of filter location occurred ($b = 9.89 \cdot 10^{-2}, SE = 8.53 \cdot 10^{-4}, t = 115.88$), because saccade amplitudes increased with central filtering and decreased with peripheral filtering. A main effect of filter type ($b = 1.17 \cdot 10^{-2}, SE = 2.34 \cdot 10^{-3}, t = 5.00$) reflected longer saccades with low-pass filtering than with high-pass filtering. As indicated by an interaction of filter type and filter location ($b = 2.01 \cdot 10^{-2}, SE = 8.53 \cdot 10^{-4}, t = 23.49$), this difference between filter types was stronger with central filtering than with peripheral filtering. Saccade amplitudes were similar with peripheral low-pass and high-pass filters of small or medium size. As in Experiment 1, completely low-pass filtered scenes provoked longer saccade amplitudes than completely high-pass filtered scenes ($b = 4.51 \cdot 10^{-2}, SE = 3.10 \cdot 10^{-3}, t = 14.56$).

Main effects of filter size indicated increasing mean saccade amplitudes as filters became larger, that is, longer amplitudes with medium and large filters than with small filters ($b = 1.02 \cdot 10^{-2}, SE = 8.98 \cdot 10^{-4}, t = 11.37$), and longer amplitudes with medium than with large filters ($b = 4.37 \cdot 10^{-3}, SE = 1.05 \cdot 10^{-3}, t = 4.15$). However, these main effects were largely due to the effects of central filtering on saccade amplitudes and were therefore qualified by interactions of filter size and filter location ($b = 3.15 \cdot 10^{-2}, SE = 8.98 \cdot 10^{-4}, t = 35.09$ for small versus medium and large filters, and $b = 9.51 \cdot 10^{-3}, SE = 1.05 \cdot 10^{-3}, t = 9.03$ for medium versus large filters). The increase of saccade amplitudes with increasing central filter size was more pronounced with low-pass filters than with high-pass filters, as indicated by interactions of filter type and filter size ($b = 1.02 \cdot 10^{-2}, SE = 8.98 \cdot 10^{-4}, t = 11.32$ for small versus medium and large filters, and $b = 6.57 \cdot 10^{-3}, SE = 1.05 \cdot 10^{-3}, t = 6.24$ for medium versus large filters).

Interestingly, mean saccade amplitudes were hardly affected by filter size with peripheral filtering, which is in contrast to previous studies demonstrating a decrease of mean amplitudes with increasing low-pass filter size (Nuthmann, 2013, 2014). Inspection of the distributions of saccade amplitudes for peripheral low-pass and high-pass filtering (Figures 4.4c and 4.4d respectively) sheds light on the absent effect of filter size. In the filter conditions, the modes of the distributions mostly shifted toward the respective radius of the gaze-contingent window, and the amount of short saccades increased with increasing filter size. Thus, viewers programmed more saccades inside or to the border of the unfiltered central region with increasing filter size. With low-pass filtering, this effect was counteracted by an increased number of long saccades from small to large filters, causing the effect of decreasing mean saccade amplitudes with increasing filter size to disappear. The effects might reflect a viewing strategy where smaller parts of the scene are inspected through a series of short saccades inside the unfiltered central region, followed by long saccades to target new parts of the scene. All variance components and fixed effects of the LMM can be found in the Appendix (Table 4.4).

4.4.3 Task performance

Mean proportion of correct answers to the memory questions are illustrated in Figure 4.3b. Compared with unfiltered scene viewing, task performance decreased when the scene was partly or entirely filtered ($b = -2.89 \cdot 10^{-1}, SE = 9.06 \cdot 10^{-2}, z = -3.19, p = .001$), with the decrease being stronger when the entire scene was filtered ($b = -6.33 \cdot 10^{-1}, SE = 8.57 \cdot 10^{-2}, z = -7.38, p < .001$). Furthermore, task performance decreased with increasing filter size, as shown by two main effects of filter size ($b = -2.87 \cdot 10^{-1}, SE = 7.24 \cdot 10^{-2}, z = -3.96, p < .001$ for small versus medium and large filters, and $b = -1.78 \cdot 10^{-1}, SE = 8.13 \cdot 10^{-2}, z = -2.19, p = .029$ for medium versus large filters). There were no other significant effects.

4.4.4 Summary

Spatial-frequency filtering with varying filter size clearly modulated eye-movement behavior. Task performance decreased when the scene was filtered and decreased more when filters became larger. Fixation durations increased in most filter conditions. Central high-pass and peripheral low-pass filtering elicited longer fixation durations than central low-pass and peripheral high-pass filtering, suggesting that more processing time was invested when the available spatial frequencies matched the viewers' tasks (fovea: analysis of details, periphery: saccade target selection). Increasing filter size prolonged fixation durations monotonically with peripheral filtering, but had no effect on fixation durations with central filtering. The latter implies that fixation durations were mainly controlled by processing difficulty in the fovea.

Saccade amplitudes largely adapted to increasing processing difficulty. Amplitudes increased with central filtering and decreased with peripheral filtering. With medium and large filters, amplitudes deviated more from unfiltered viewing when spatial frequencies were potentially less useful for processing. Furthermore, increasing filter size provoked more long saccades with central filtering and more short saccades with peripheral filtering.

Effects of filter size were pronounced with either mean fixation durations or saccade amplitudes, but not both. With peripheral filtering, filter size had clear effects on mean fixation durations, but not on mean saccade amplitudes; with central filtering, the pattern was reversed. This suggests a trade-off between temporal and spatial aspects of eye-movement behavior.

4.5 Discussion

The present study investigated how the availability of different spatial frequencies in the central or the peripheral visual field affects eye-movement behavior during real-world scene viewing. For this purpose, high-pass or low-pass filters were applied either inside or outside a gaze-contingent window while the respective other region remained unfiltered. Filter level was varied in Experiment 1, and filter size was varied in Experiment 2. Results demonstrate that temporal as well as spatial aspects of eye-movement behavior are modulated by the type, size, and level of filtering.

4.5.1 Fixation durations do not always increase with processing difficulty

Of particular interest was the question how fixation durations adapt to varying processing difficulty due to spatial-frequency filtering. Generally, fixation durations are assumed to increase as visual-cognitive processing difficulty increases. Accordingly, prior research reported increased fixation durations when low spatial frequencies are filtered from the central or the peripheral visual field during scene viewing (Loschky & McConkie, 2002; Loschky et al., 2005; Nuthmann, 2013, 2014; Parkhurst et al., 2000; van Diepen & Wampers, 1998), with the effects getting stronger with increasing filter size (Loschky & McConkie, 2002; Nuthmann, 2013, 2014; Parkhurst et al., 2000).

In contrast to these findings, the present results indicate that fixation durations increase with higher processing difficulty due to spatial-frequency filtering only when difficulty is moderately increased. The increase of fixation durations was more pronounced with central high-pass and peripheral low-pass filtering than with central low-pass and peripheral high-pass filtering. This result replicates our findings from previous studies (Cajar, Schneeweiß, et al., 2016; Laubrock et al., 2013) and suggests that viewers invest more time for processing spatial frequencies that are potentially more useful for foveal analysis of details (high spatial frequencies) and peripheral target selection (low spatial frequencies). When processing becomes too difficult and information cannot be efficiently used for gaze control, fixation durations do not prolong any further. In this sense, viewing behavior appears to be quite economical.

Most impressively, this behavior is reflected in the effects of filter level and filter size. First, fixation durations were unaffected by increasing filter level with peripheral high-pass filtering. High spatial frequencies cannot be resolved in the low-resolution periphery and are therefore hardly useful for saccade target selection. Thus, increasing fixation durations with increasing level of high-pass filtering was probably not worthwhile. Increasing the level of peripheral low-pass filtering, on the other hand, prolonged fixation durations, indicating that strong low-pass filtered information can still be used for peripheral target selection. This result contradicts Loschky and colleagues (Loschky & McConkie, 2002; Loschky et al., 2005), who found that the level of detectable peripheral low-pass filtering hardly affected fixation durations. Second, fixation durations were unaffected by increasing filter size with both central high-pass and low-pass filtering. Compared with unfiltered scene viewing, fixation durations increased with spatial-frequency filtering in the foveal visual field, but additional filtering of the parafovea did not prolong fixation durations any further (contrary to Nuthmann, 2014). Thus, fixation durations appeared to be mainly controlled by processing difficulty in the fovea. Increasing filter size in the peripheral visual field, though, increased fixation durations monotonically (replicating Loschky & McConkie, 2002; Nuthmann, 2013, 2014; Parkhurst et al., 2000). Still, fixation durations were similar to or shorter than the unfiltered control condition with medium and large peripheral high-pass filters respectively. Presumably, foveal and parafoveal scene information could be processed rather efficiently in these conditions, but little time was invested for saccade target selection in an apparently uniform gray periphery where the extraction of scene information was rather difficult.

We conclude that fixation durations prolong with increasing processing difficulty due to the type, level, and size of spatial-frequency filters as long as the available information can be analyzed in a reasonable amount of time and is critical for gaze control; otherwise, stimulus-independent timing takes over.

4.5.2 Peripheral information is critical for the control of fixation duration

Interestingly, fixation durations were affected more strongly by increasing processing difficulty in the peripheral visual field than in the central visual field. First, mean differences in fixation durations between filter types were notably larger with peripheral filtering in both experiments, suggesting that the kind of available spatial-frequency information is more important in the peripheral than in the central visual field. This seems reasonable—since the fovea and parafovea are sensitive to a larger band of spatial frequencies than the periphery, they cope better with missing frequencies when processing the scene.

Second, foveal analysis was less affected by increasing processing difficulty due to increasing filter size and filter level than peripheral selection. With both low-pass and high-pass filtering, fixation durations increased with central filtering, but were not affected by the size of the central filter and increased little as filter level became stronger. Based on these results, we conclude that (i) the fovea is critical for the control of fixation durations whereas the parafovea plays a subordinate role, and (ii) the level of scene degradation via spatial filtering of the foveal stimulus has little effect on fixation durations. With peripheral filtering, however, filter level and filter size had considerable effects on fixation durations. Mean durations increased with both filter types as the filtered part of the scene became larger, thus making saccade target selection increasingly difficult. With peripheral low-pass filtering, fixation durations also increased with increasing level of filtering.

The present findings suggest that peripheral vision is more sensitive to the kind of available spatial-frequency information than central vision. Contrary to the common notion that fixation durations are dominated by foveal processing, our results show that fixation durations can be controlled considerably by peripheral processing.

4.5.3 Switching costs between filtered and unfiltered scene regions

Peripheral low-pass filtering with moderate and strong filters (Experiment 1) and with large filters (Experiment 2) provoked mean fixation durations that were longer than the mean duration in the filtered control condition in which the entire scene was low-pass filtered. Thus, gaze-contingent filtering entailed switching costs between the filtered peripheral and the unfiltered central scene region when the filter was strong or large enough. Reingold and Loschky (2002) found similar effects, with longer saccade latencies to a peripheral target when only the peripheral visual field was low-pass filtered compared to low-pass filtering of the entire scene. The authors reasoned that low-pass filtering of only part of the visual field increases the saliency of the unfiltered region; as a consequence, the competition for attention between the filtered and unfiltered scene region might increase and therefore prolong fixation durations compared to viewing entirely filtered scenes.

4.5.4 Saccade amplitudes adapt to processing difficulty

Saccadic behavior strongly adapted to the manipulated parameters of the stimulus. Compared to viewing unfiltered scenes, mean saccade amplitudes lengthened with central filtering and shortened with peripheral filtering. This effect replicates previous findings (Cajar, Schneeweiß, et al., 2016; Foulsham et al., 2011; Laubrock et al., 2013; Loschky & McConkie, 2002; Nuthmann, 2013, 2014; Reingold & Loschky, 2002; van Diepen & Wampers, 1998) and indicates a preference for unfiltered scene regions as saccade targets—central filtering elicits a bias for targeting scene regions farther away from the current fixation position, whereas peripheral filtering elicits more focussed gaze patterns in the unfiltered central region. This opposing viewing behavior with central and peripheral filtering strengthened as scene processing became more difficult. First, saccades increasingly lengthened with central filtering and increasingly shortened with peripheral filtering as filter level and filter size increased. Second, saccade amplitudes increased more when the central visual field was low-pass filtered than when it was high-pass filtered, whereas they decreased more when the peripheral visual field was high-pass filtered than when it was low-pass filtered.

In summary, saccade target selection was progressively impaired by increasing processing difficulty due to filtering. As shown by Cajar, Schneeweiß, et al. (2016), these modulations of saccadic selection reflect modulations of attention, with smaller saccades indicating a withdrawal of attention from the periphery (i.e., tunnel vision) and larger saccades indicating an attentional bias toward the periphery.

4.5.5 Trade-off between fixation duration and saccade amplitude

Increasing filter size in the peripheral visual field increased mean fixation durations but hardly affected mean saccade amplitudes; increasing filter size in the central visual field, on the other hand, increased saccade amplitudes but left fixation durations unaffected. Varying filter size therefore provoked a viewing behavior with a trade-off between fixation durations and saccade amplitudes (see Jacobs, 1986), with opposing effects for central and peripheral filtering.

This trade-off was also evident with increasing processing difficulty due to filter level—as filter level increased with peripheral high-pass filtering, fixation durations were unaffected, but saccade amplitudes increasingly shortened. Furthermore, mean fixation durations were similar to or even below the mean of the unfiltered control condition with medium and large peripheral high-pass filters, whereas saccade amplitudes shortened. We provide the following explanation for these effects. High-pass filtering inherently attenuates luminance, contrast, and color information of the stimulus. Therefore, the segregation of peripheral objects from their background becomes harder with increasing filter level; with a strong high-pass filter, the peripheral scene resembles a uniform gray background. Thus, selecting saccade targets from strongly high-pass filtered peripheral scene regions was probably too difficult, causing saccade targets to be chosen either from the unfiltered central part of the scene (hence shorter saccade amplitudes) or less carefully from the filtered periphery. Both strategies can reduce the time needed to select the next saccade target and thus shorten mean fixation durations as observed. However, when the peripheral filter gets so large that it approaches foveal vision, the analysis of both the peripheral and the central stimulus is affected, which might explain the observed increase of mean fixation duration with large peripheral high-pass filters.

In summary, increasing filter level and size often modulated either mean fixation durations or saccade amplitudes. Thus, moderate attenuation of spatial frequencies affects temporal and spatial aspects of eye-movement behavior, but with a further increase of processing difficulty resources are often preserved by adapting either saccade timing or saccadic selection, but not both.

4.6 Conclusions

The present study demonstrates that eye-movement behavior during real-world scene viewing is impaired with increasing processing difficulty due to spatial-frequency filters of varying type, size, and level in the central or the peripheral visual field. The adaptation to increasing processing difficulty is particularly evident in the modulation of saccade target selection (i.e., saccade amplitudes). Saccade timing is not necessarily adapted to processing difficulty—when visual-cognitive processing becomes too time-consuming, fixation durations are not prolonged any further, but instead, stimulus-independent timing is adapted. This finding is in good agreement with the assumption of a psychologically plausible trade-off relation between visual processing time and saccade generation.

4.7 Acknowledgments

This work was funded by Deutsche Forschungsgemeinschaft (grants LA 2884/1 to J.L. and EN 471/10 to R.E.). We thank Heiko Schütt for helpful comments and Petra Schienmann and our student assistants for their help during data collection.

4.8 Appendix

4.8.1 Variance components and fixed effects estimated with linear mixed-effects models

Table 4.1: Variance components and fixed effects for fixation durations in Experiment 1. *t*-values exceeding an absolute value of 1.96 were regarded as being statistically significant on the two-tailed 5% level.

Variance components	SD ($\cdot 10^{-2}$)
Participants	9.48
Scenes	2.44
Residuals	41.44

Fixed effects	Coefficient ($\cdot 10^{-2}$)	SE ($\cdot 10^{-3}$)	t
Unfiltered control low-pass vs. high-pass	2.00	5.41	3.7
Unfiltered controls vs. all filter conditions	4.37	2.36	18.5
Filtered control low-pass vs. high-pass	2.35	5.51	4.3
Filtered controls vs. gaze-contingent filters	1.12	2.44	4.6
Filter type	2.29	3.64	6.3
Filter location	1.56	1.84	8.5
Weak vs. moderate and strong filters (W-MS)	2.20	1.94	11.3
Moderate vs. strong filter (M-S)	0.90	2.26	4.0
Filter type × Filter location	2.79	1.84	15.2
Filter type × W-MS	1.83	1.94	9.4
Filter type × M-S	0.84	2.26	3.7
Filter location × W-MS	0.87	1.94	4.5
Filter location × M-S	-0.03	2.26	-0.1
Filter type × Filter location × W-MS	1.34	1.94	6.9
Filter type × Filter location × M-S	0.57	2.25	2.5

Table 4.2: Variance components and fixed effects for saccade amplitudes in Experiment 1. t-values exceeding an absolute value of 1.96 were regarded as being statistically significant on the two-tailed 5% level.

Variance components	SD ($\cdot 10^{-2}$)
Participants	5.57
Scenes	2.39
Residuals	25.45

Fixed effects	Coefficient ($\cdot 10^{-2}$)	SE ($\cdot 10^{-3}$)	t
Unfiltered control low-pass vs. high-pass	-1.19	4.07	-2.92
Unfiltered controls vs. all filter conditions	0.16	1.42	1.11
Filtered control low-pass vs. high-pass	2.36	4.12	5.73
Filtered controls vs. gaze-contingent filters	1.34	1.47	9.08
Filter type	1.31	3.28	3.98
Filter location	8.14	1.11	73.54
Weak vs. moderate and strong filters (W-MS)	0.007	1.17	0.06
Moderate vs. strong filter (M-S)	0.27	1.36	1.96
Filter type × Filter location	1.01	1.11	9.13
Filter type × W-MS	-0.65	1.17	-5.53
Filter type × M-S	-0.06	1.36	-0.47
Filter location × W-MS	-3.67	1.17	-31.33
Filter location × M-S	-2.65	1.36	-19.48
Filter type × Filter location × W-MS	-1.01	1.17	-8.62
Filter type × Filter location × M-S	-0.74	1.36	-5.45

Table 4.3: Variance components and fixed effects for fixation durations in Experiment 2. *t*-values exceeding an absolute value of 1.96 were regarded as being statistically significant on the two-tailed 5% level.

Variance components	SD ($\cdot 10^{-2}$)
Participants	11.72
Scenes	2.30
Residuals	42.21

Fixed effects	Coefficient ($\cdot 10^{-2}$)	SE ($\cdot 10^{-3}$)	t
Unfiltered control low-pass vs. high-pass	1.30	5.51	2.36
Unfiltered controls vs. all filter conditions	5.41	2.49	21.76
Filtered control low-pass vs. high-pass	1.91	5.78	3.31
Filtered controls vs. gaze-contingent filters	3.21	2.66	12.09
Filter type	2.99	3.54	8.44
Filter location	1.68	1.92	8.74
Small vs. medium and large filters (S-ML)	3.77	2.02	18.64
Medium vs. large filter (M-L)	4.69	2.37	19.77
Filter type × Filter location	5.37	1.92	27.96
Filter type × S-ML	0.36	2.02	1.80
Filter type × M-L	0.78	2.37	3.30
Filter location × S-ML	4.66	2.02	23.05
Filter location × M-L	4.42	2.37	18.66
Filter type × Filter location × S-ML	0.76	2.02	3.78
Filter type × Filter location × M-L	0.25	2.37	1.04

Table 4.4: Variance components and fixed effects for saccade amplitudes in Experiment 2. t-values exceeding an absolute value of 1.96 were regarded as being statistically significant on the two-tailed 5% level.

Variance components	SD ($\cdot 10^{-2}$)
Participants	2.66
Scenes	1.69
Residuals	19.12

Fixed effects	Coefficient ($\cdot 10^{-2}$)	SE ($\cdot 10^{-3}$)	t
Unfiltered control low-pass vs. high-pass	0.89	3.00	2.96
Unfiltered controls vs. all filter conditions	0.39	1.10	3.53
Filtered control low-pass vs. high-pass	4.51	3.10	14.56
Filtered controls vs. gaze-contingent filters	1.37	1.18	11.60
Filter type	1.17	2.34	5.00
Filter location	9.89	0.85	115.88
Small vs. medium and large filters (S-ML)	1.02	0.90	11.37
Medium vs. large filter (M-L)	0.44	1.05	4.15
Filter type × Filter location	2.01	0.85	23.49
Filter type × S-ML	1.02	0.90	11.32
Filter type × M-L	0.66	1.05	6.24
Filter location × S-ML	3.15	0.90	35.09
Filter location × M-L	0.95	1.05	9.03
Filter type × Filter location × S-ML	0.86	0.90	9.59
Filter type × Filter location × M-L	-0.002	1.05	-0.02

Chapter 5

General summary and conclusions

Eye movements serve as a window into ongoing visual-cognitive processes and can thus be used as a tool to understand how people perceive and process real-world scenes. As in any other domain of eye-movement research, a key issue for understanding eye-movement control during scene viewing is the roles played by central and peripheral vision. Yet, rather little is known about central and peripheral processing during scene viewing and how they are coordinated within a fixation. Furthermore, until recently, research on gaze control during scene viewing has mainly been focused on the question *where* people move their eyes in scenes, whereas the factors determining *when* people move their eyes have long been neglected. This disparity is reflected in a vast number of computational models on fixation locations in scenes (for reviews, see Borji & Itti, 2013; Tatler et al., 2011) opposing only one computational model for fixation durations (Nuthmann et al., 2010).

The present thesis therefore investigated how central and peripheral vision contribute to spatial and, particularly, to temporal aspects of eye-movement control during scene viewing. We conducted a series of experiments with selective degradation of scene information in the central or the peripheral visual field and developed a computational model on how foveal and peripheral processing might be coordinated for the control of fixation duration. The following sections discuss the roles of central and peripheral vision for the control of fixation duration, the distribution of attention, and implications for computational models of eye-movement control during scene viewing.

5.1 Control of fixation duration

New information from the scene can only be acquired during fixations. Thus, fixation duration is one of the most important measures of visual-cognitive processing (Rayner, 1998). To investigate the contributions of central and peripheral vision to variations in fixation duration, we degraded scene information in the central or the peripheral visual field by spatial-frequency filtering, thus hampering detailed foveal analysis or peripheral target selection respectively.

5.1.1 Fixation duration does not always increase with processing difficulty

It is a general assumption that fixation durations adapt to visual-cognitive processing difficulty (Rayner, 1998). In line with this notion, fixation durations during scene viewing have usually been shown to prolong with increasing processing difficulty of the stimulus. Consequently, filtering selective spatial frequencies in the central or the peripheral visual field in the present experiments was expected to increase fixation durations to the extent that the missing information in the respective region is important for processing that stimulus. However, the results paint a different picture. In all experiments, fixation durations increased considerably when available spatial frequencies were potentially more useful for the task at hand, that is, matching the sensitivity of the central visual field to higher spatial frequencies for detailed foveal analysis (central high-pass filtering) and the sensitivity of the peripheral visual field to low spatial frequencies for saccade target selection (peripheral low-pass filtering). When spatial frequencies did not match this specialization of central and peripheral vision (i.e., with central low-pass and peripheral high-pass filtering), fixation durations were mostly similar to the durations with unfiltered scene viewing or even shorter. Furthermore, chapter 4 showed that increasing filter level and filter size in the central visual field as well as increasing filter level with peripheral high-pass filtering had little effect on fixation durations.

Thus, all experiments consistently demonstrate that fixation duration does not necessarily prolong with increasing processing difficulty of the stimulus due to increasing filter size, filter level, or non-matching filter type. Fixations only prolonged when difficulty was moderately increased and the investment of more processing time was worthwhile, that is, when a higher amount of useful information was anticipated. When processing became too difficult and time-consuming because the available information could not be used efficiently for gaze control, fixation durations did not increase, but were rather stimulus-independent or even decreased. This pattern emerged both with increasing processing difficulty in the central and the peripheral visual field, where filtering hampered either foveal analysis or saccade target selection, and indicates economical behavior with a trade-off between processing time and saccade generation.

This trade-off is also reflected by the adjustment of saccade target selection under conditions of high processing difficulty when fixation durations are not adapted. In all experiments, amplitudes increasingly shortened with peripheral filtering and lengthened with central filtering as filters became larger, stronger, or attenuated more useful frequencies. This pattern shows that viewers avoided saccades to filtered scene regions and increasingly adapted saccadic behavior as processing difficulty increased. Thus, viewers seem to balance the costs of increasing fixation duration and adjusting saccadic behavior.

We believe that the present results on fixation durations provide evidence against a direct control of fixation duration by the central or the peripheral stimulus. Actually, the paradigm used in the present experiments does not allow to distinguish between local control theories of fixation duration—because filters are blocked across trials, it cannot be determined whether each individual fixation adapts immediately to processing demands (direct control), or whether fixation duration is globally adjusted to task difficulty (indirect control), for example, at the beginning of each trial. Yet, the direct control theory assumes that fixation duration adjusts to current processing demands and increases with the difficulty of the currently fixated stimulus (Rayner & Pollatsek, 1981). Under this assumption, each individual fixation, and consequently also mean fixation durations should have increased under conditions of high processing difficulty. However, the results often revealed no adjustment of fixation duration to increasing processing difficulty in the central or the peripheral visual field, or even a decrease of fixation durations; in contrast, fixation durations increased when processing difficulty was moderately increased.

It is left for future scene perception research to determine more directly to what degree fixation duration is controlled immediately by the difficulty of the stimulus currently present in central or peripheral vision. The present paradigm of central or peripheral spatial-frequency filtering might be adapted for this purpose by changing central or peripheral frequency filters during *individual* fixations within the trial (compared with using the same filter location and filter type during the entire trial).

5.1.2 The role of peripheral vision for the control of fixation duration

Previous research suggests that fixation duration is mainly controlled by foveal information processing and only affected weakly by peripheral processing (Foulsham et al., 2011; Henderson, 1992; Henderson & Ferreira, 1993; Hooge & Erkelens, 1999). Furthermore, peripheral information is considered to be rather unimportant at the beginning of the fixation and only to be processed later within the fixation after the foveal stimulus has been processed to some degree (Henderson, 1992; Rayner, 2009; van Diepen & Wampers, 1998; van Diepen & d'Ydewalle, 2003).

Contrary to these notions, both the present experiments and the computational model suggest that the processing of peripheral information plays a decisive role for the control of fixation duration during scene viewing. In all experiments, peripheral scene degradation (especially low-pass filtering) had considerable effects on fixation duration, often stronger than and different from the effects of central scene degradation (see chapter 4 in particular). We developed a computational model that accounted for the individual effects of central and peripheral processing difficulty on fixation duration and assigned a key role to peripheral vision for the control of fixation duration.

First, the model assumes a mixed control of fixation durations. A random timer generates fixation durations autonomously with a preferred mean rate, but the timer can be inhibited at any point in time by ongoing visual-cognitive processing demands. Thus, in general, saccades are generated independently from stimulus processing, but current processing demands can inhibit saccade generation. Second, spatial visual-cognitive processing is accounted for by activations in two spatial compartments, a foveal and a peripheral compartment. Activations in these two compartments evolve in parallel and independently from the beginning of the fixation; yet, foveal and peripheral processing can interact in order to inhibit the saccade timer and thus to prolong fixation durations.

Experimental evidence for parallel and independent processing of foveal and peripheral information from the beginning of the fixation was given subsequently for a simple task environment with Gabor patches on a uniform background (Ludwig et al., 2014). Our computational model can also explain eye-movement behavior with parallel and independent processing in foveal and peripheral vision in a more complex task environment like real-world scene viewing. Numerical simulations with the model reproduced the specific effects of central and peripheral scene degradation on fixation durations presented in chapter 2 and revealed the interaction of foveal and peripheral processing as the key mechanism for generating long fixation durations.

In summary, we assign a more prominent role to peripheral vision for the control of fixation duration than suggested by previous research. Proposing only the second model on the control of fixation durations during scene viewing (following Nuthmann et al., 2010) with the novel assumption on spatially distributed processing, the present work not only provides new experimental findings, but also makes an important theoretical contribution on how fixation durations might be controlled during scene viewing.

5.1.3 Further outlook: Time course of central and peripheral processing

A question only little is known about is the time course of central and peripheral processing during scene viewing, that is, how central and peripheral processing are coordinated within a fixation. Do they evolve in parallel or sequentially? If they evolve (at least partly) sequentially, when exactly does processing switch between central and peripheral processing? These questions will probably be difficult to answer for complex and highly variable stimuli like real-world scenes (Rayner, 2009). As aforementioned, it is a widespread assumption that foveal information is processed some time before peripheral information (Henderson, 1992; Rayner, 2009; van Diepen & Wampers, 1998; van Diepen & d'Ydewalle, 2003).

In the present work we implemented a computational model proposing parallel and independent processing of foveal and peripheral scene information from the beginning of the fixation with continuous crosstalk between these processing streams. Future studies are required that also investigate *experimentally* how time within a fixation is allocated to central and peripheral processing during scene viewing. As a first step, onset delay experiments masking only central or peripheral scene information for varying time periods during individual fixations might be used (see, e.g., van Diepen & Wampers, 1998; van Diepen & d'Ydewalle, 2003, for central or peripheral scene onset delays within every single fixation). Fixation durations should be affected by central or peripheral masking to the extent that the respective information is used during this time of fixation.

5.2 Distribution of attention

Numerous studies have reported a characteristic pattern of mean saccade amplitudes with central and peripheral scene degradation compared with undegraded scene viewing: longer saccade amplitudes with central degradation and shorter saccade amplitudes with peripheral degradation. This pattern indicates that viewers avoid scene regions of increased processing difficulty as saccade targets.

All experiments presented in chapters 2 to 4 replicated this amplitude pattern. In chapter 3, we investigated whether this pattern of saccade amplitudes observed with increased central and peripheral processing difficulty reflects corresponding modulations of visual attention. As saccades and attention are tightly coupled (Deubel & Schneider, 1996; Hoffman & Subramaniam, 1995; Kowler et al., 1995), saccade amplitudes are usually considered to reflect the spatial extent of attention. Under this assumption, smaller saccade amplitudes during peripheral scene degradation should reflect a shrinkage of the attentional focus (i.e., tunnel vision) whereas longer saccade amplitudes during central scene degradation should reflect an attentional bias toward the periphery. Our results support this assumption, as viewers were less sensitive to peripheral target stimuli with peripheral degradation and less sensitive to central target stimuli with central degradation. These findings indeed suggest an attentional bias toward the peripheral visual field when central processing difficulty was increased and an attentional bias toward the central visual field when peripheral processing difficulty was increased. Thus, viewers partly withdraw attention from scene regions in which the extraction of information is difficult. This withdrawal of attention entails corresponding changes in saccadic behavior.

One observation from chapter 3 deserving further consideration is that central scene degradation did not affect the ability to detect peripheral targets. This finding is in conflict with previous research showing that performance in peripheral detection tasks decreases with increased foveal load (e.g. Crundall et al., 1999; Ikeda & Takeuchi, 1975; Williams, 1985, 1989, 1995). These findings are assumed to reflect tunnel vision where attentional resources are withdrawn from the periphery due to increased central processing demands. However, similar to our results, Ludwig et al. (2014) also found that peripheral selection was unaffected by foveal load. Both our experiments as well as those by Ludwig et al. (2014) manipulated low-level visual features (spatial frequency and luminance contrast or orientation respectively), compared to the higher-level cognitive tasks (e.g., letter discrimination) by studies reporting tunnel vision with increased foveal load. Thus, it seems that foveal load might have different effects on attention depending on the level of processing required by the task.

Finally, another interesting finding came from our post-hoc analyses on the correlation between target detection probability and the amplitude and direction of the saccade following target onset. In both experiments presented in chapter 3, target detection probability was highest when saccade amplitudes approximated the eccentricity of the target, irrespective of saccade direction. This finding suggests that attention was allocated in a ring-shaped fashion during scene viewing, as previously shown for letter identification tasks in simple environments (Egly & Homa, 1984; Juola et al., 1991). However, as these results were only obtained post hoc, future studies need to test systematically for the possibility of a ring-shaped allocation of attention during scene viewing.

5.3 Implications for models of eye-movement control during scene viewing

The experimental results presented in chapters 2 to 4 suggest that (i) effects of scene degradation on fixation durations cannot be explained by the direct control account, (ii) peripheral vision plays a critical role for the control of fixation duration, and (iii) scene degradation has different effects on eye-movement parameters when applied to central or peripheral vision. To attain a deeper understanding of the underlying processes, these findings were implemented in a new computational model for the control of fixation duration during scene viewing.

Nearly all computational models of eye-movement control during scene viewing ignore the different roles of central and peripheral vision, thus assuming homogeneous processing across the visual field (but see Engbert et al., 2015; Wischnewski et al., 2010; Zelinsky, 2008). These models are therefore unable to explain any differences in central or peripheral processing as found in the present experiments. However, as Ludwig et al. (2014, p. E297) stated: "Any model of eye movement control needs to solve the same problem, regardless of the domain of application: How are foveal analysis and peripheral selection coordinated?" The computational model proposed in the present work provides the first possible answer to this question in scene viewing.

Furthermore, most models on gaze control during scene viewing predict fixation locations, but completely ignore fixation durations. A notable exception is CRISP (Nuthmann et al., 2010), the first model for the control of fixation durations during scene viewing. Yet, CRISP also disregards the inhomogeneity of the visual field and thus the different processing in central and peripheral vision. However, the present model proposes that the variance in fixation durations partly arises from the complicated interaction of foveal and peripheral processing.

In its current implementation, our model does not account for the influence of saccade programming on fixation duration. Following the experimental results of Becker and Jürgens (1979), existing computational models of eye-movement control (e.g., Engbert et al., 2005; Nuthmann et al., 2010; Reichle et al., 1998) usually implement saccade programming as a two-stage process with an initial labile stage, in which the saccade program can still be canceled, and an ensuing nonlabile stage, in which the program can no longer be canceled. The implementation of saccade programming in the present model, which evidently has a part in the control of fixation duration[1] is left for future work.

---

[1] For example, simulations with the CRISP model identified the cancelation of labile saccade programs as a critical mechanism for explaining fixation durations in scene and mask onset delay experiments (Nuthmann et al., 2010).

Finally, in order to fully understand the distribution of attention and visual-cognitive processing during scene viewing, fixation locations need to be weighted by fixation durations (Henderson, 2003) and spatial interactions in gaze patterns have to be taken into account (Engbert et al., 2015). Thus, a complete and plausible model of eye-movement control during scene viewing has to consolidate spatial and temporal aspects of gaze control. To this end, the present computational model for fixation durations could be incorporated with an existing model predicting fixation locations and attentional dynamics (e.g., Engbert et al., 2015), thus taking the next step toward a full model of eye-movement control during scene viewing.

5.4 Conclusions

The present thesis investigated how central and peripheral vision contribute to eye-movement control during scene viewing, with a particular focus on the control of fixation durations. For all experimental manipulations, we found significant differences between central and peripheral vision both for temporal and spatial aspects of eye-movement control. We showed that saccadic selection increasingly adapts to central and peripheral processing difficulty and reflects corresponding modulations of visual attention, whereas fixation durations do not necessarily prolong with increasing difficulty of the central or the peripheral stimulus. Our experiments and computational model suggest a decisive role for peripheral processing in the control of fixation duration. Furthermore, the proposed model is the first computational model on fixation durations taking separate processing in central and peripheral vision into account; it suggests that fixation durations during scene viewing are partly controlled by the interaction of central and peripheral processing. The present work highlights the importance of accounting for different processing in central and peripheral vision when studying scene perception and entails new implications for experimental research and computational models in this area.

References

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412. doi:10.1016/j.jml.2007.12.005.

Baddeley, R. J., & Tatler, B. W. (2006). High frequency edges (but not contrast) predict where we fixate: A Bayesian system identification analysis. *Vision Research*, *46*, 2824–2833. doi:10.1016/j.visres.2006.02.024.

Banks, M. S., Sekuler, A. B., & Anderson, S. J. (1991). Peripheral spatial vision: Limits imposed by optics, photoreceptors, and receptor pooling. *Journal of the Optical Society of America A*, *8*(11), 1775–1787. doi:10.1364/JOSAA.8.001775.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, R. H. (2015). Parsimonious mixed models. Retrieved from http://arxiv.org/abs/1506.04967.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi:10.18637/jss.v067.i01.

Becker, W., & Jürgens, R. (1979). An analysis of the saccadic system by means of double step stimuli. *Vision Research*, *19*, 967–983. doi:10.1016/0042-6989(79)90222-0.

Bertera, J. H., & Rayner, K. (2000). Eye movements and the span of the effective stimulus in visual search. *Perception & Psychophysics*, *62*(3), 576–585. doi:10.3758/BF03212109.

Borji, A., & Itti, L. (2013). State-of-the-art in visual attention modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(1), 185–207. doi:10.1109/TPAMI.2012.89.

Boucart, M., Lenoble, Q., Quettelart, J., Szaffarczyk, S., Despretz, P., & Thorpe, S. J. (2016). Finding faces, animals, and vehicles in far peripheral vision. *Journal of Vision*, *16(2):10*, 1–13. doi:10.1167/16.2.10.

Boucart, M., Moroni, C., Thibaut, M., Szaffarczyk, S., & Greene, M. (2013). Scene categorization at large visual eccentricities. *Vision Research*, *86*, 35–42. doi:10.1016/j.visres.2013.04.006.

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations.

Journal of the Royal Statistical Society, *26B*, 211–252. Retrieved from http://www.jstor.org/stable/2984418.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436. doi:10.1163/156856897X00357.

Buswell, G. T. (1935). *How people look at pictures.* Chicago: University of Chicago Press.

Cajar, A., Engbert, R., & Laubrock, J. (2016). Spatial frequency processing in the central and peripheral visual field during scene viewing. *Vision Research*, *127*, 186–197. doi:10.1016/j.visres.2016.05.008.

Cajar, A., Laubrock, J., & Engbert, R. (2013). Control of saccade timing by foveal and peripheral spatial frequencies in natural scenes. In K. Holmqvist, F. Mulvey, & R. Johansson (Eds.), *Book of Abstracts of the 17th European Conference on Eye Movements. Journal of Eye Movement Research, 6.* (3), 266.

Cajar, A., Schneeweiß, P., Engbert, R., & Laubrock, J. (2016). Coupling of attention and saccades when viewing scenes with central and peripheral degradation. *Journal of Vision*, *16(2):8*, 1–19. doi:10.1167/16.2.8.

Campbell, F. W., & Robson, J. G. (1968). Application of fourier analysis to the visibility of gratings. *Journal of Physiology*, *197*, 551–566. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1351748/.

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):6*, 1–15. doi:10.1167/9.3.6.

Chastain, G. (1992a). Analog versus discrete shifts of attention across the visual field. *Psychological Research*, *54*(3), 175–181. doi:10.1007/BF00922096.

Chastain, G. (1992b). Time-course of sensitivity changes as attention shifts to an unpredictable location. *The Journal of General Psychology*, *119*(2), 105-111. doi:10.1080/00221309.1992.9921164.

Cornelissen, F. W., Bruin, K. J., & Kooijman, A. C. (2005). The influence of artificial scotomas on eye movements during visual search. *Optometry and Vision Science*, *82*(1), 27–35.

Cornelissen, F. W., Peters, E. M., & Palmer, J. (2002). The Eyelink Toolbox: Eye tracking with MATLAB and the Psychophysics Toolbox. *Behavioral Research Methods, Instruments, & Computers*, *34*(4), 613–617. doi:10.3758/BF03195489.

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorial in Quantitative Methods for Psychology*, *1*(1), 71–75. Retrieved from http://tqmp.org/Content/vol01-1/p042/p042.pdf.

Crundall, D., Underwood, G., & Chapman, P. (1999). Driving experience and the functional field of view. *Perception*, *28*, 1075–1087. doi:10.1068/p2894.

Curcio, C. A., & Allen, K. A. (1990). Topography of ganglion cells in human retina. *The Journal of Comparative Neurology*, *300*, 5–25. doi:10.1002/cne.903000103.

Curcio, C. A., Sloan, K. R., Kalina, R. E., & Hendrickson, A. E. (1990). Human photoreceptor topography. *The Journal of Comparative Neurology*, *292*, 497–523. doi:10.1002/cne.902920402.

Daniel, P. M., & Whitteridge, D. (1961). The representation of the visual field on the cerebral cortex in monkeys. *The Journal of Physiology*, *159*, 203–221. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1359500/.

De Graef, P., Christiaens, D., & d'Ydewalle, G. (1990). Perceptual effects of scene context on object identification. *Psychological Research*, *52*(4), 317–329. doi:10.1007/BF00868064.

Derrington, A. M., & Lennie, P. (1984). Spatial and temporal contrast sensitivities of neurones in lateral geniculate nucleus of macaque. *The Journal of Physiology*, *357*(1), 219–240. doi:10.1113/jphysiol.1984.sp015498.

Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research*, *36*(12), 1827–1837. doi:10.1016/0042-6989(95)00294-4.

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. doi:10.1016/0042-6989(82)90113-4.

De Valois, R. L., & De Valois, K. K. (1990). *Spatial vision*. Oxford: Oxford University Press.

Dumoulin, S. O., & Wandell, B. A. (2008). Population receptive field estimates in human visual cortex. *Neuroimage*, *39*, 647–660. doi:10.1016/j.neuroimage.2007.09.034.

Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General*, *113*(4), 501–517. doi:10.1037/0096-3445.113.4.501.

Egly, R., Driver, J., & Rafal, R. D. (1994). Shifting visual attention between objects and locations: Evidence from normal and parietal lesion subjects. *Journal of Experimental Psychology: General*, *123*(2), 161–177. doi:10.1037/0096-3445.123.2.161.

Egly, R., & Homa, D. (1984). Sensitization of the visual field. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(6), 778–793. doi:10.1037/0096-1523.10.6.778.

Engbert, R., & Kliegl, R. (2001). Mathematical models of eye movements in reading: a possible role for autonomous saccades. *Biological Cybernetics*, *85*, 77–87.

doi:10.1007/PL00008001.

Engbert, R., & Kliegl, R. (2003). Microsaccades uncover the orientation of covert attention. *Vision Research*, *43*, 1035–1045. doi:10.1016/S0042-6989(03)00084-1.

Engbert, R., & Mergenthaler, K. (2006). Microsaccades are triggered by low level retinal image slip. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(18), 7192–7197. doi:10.1073/pnas.0509557103.

Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review*, *112*(4), 777–813. doi:10.1037/0033-295X.112.4.777.

Engbert, R., Trukenbrod, H. A., Barthelmé, S., & Wichmann, F. A. (2015). Spatial statistics and attentional dynamics in scene viewing. *Journal of Vision*, *15(1):14*, 1–17. doi:10.1167/15.1.14.

Eriksen, C. W., & James, J. D. S. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception & Psychophysics*, *40*(4), 225–240. doi:10.3758/BF03211502.

Eriksen, C. W., & Webb, J. M. (1989). Shifting of attentional focus within and about a visual display. *Perception & Psychophysics*, *45*(2), 175–183. doi:10.3758/BF03208052.

Eriksen, C. W., & Yeh, Y. (1985). Allocation of attention in the visual field. *Journal of Experimental Psychology: Human Perception and Performance*, *11*(5), 583–597. doi:10.1037/0096-1523.11.5.583.

Findlay, J. M. (2009). Saccadic eye movement programming: sensory and attentional factors. *Psychological Research*, *73*(2), 127–135. doi:10.1007/s00426-008-0201-3.

Findlay, J. M., & Gilchrist, I. D. (2003). *Active vision: The psychology of looking and seeing*. New York: Oxford University Press.

Findlay, J. M., & Walker, R. (1999). A model of saccade generation based on parallel processing and competitive inhibition. *Behavioral and Brain Sciences*, *22*(4), 661–721. doi:10.1017/S0140525X99002150.

Fortenbaugh, F. C., Hicks, J. C., Hao, L., & Turano, K. A. (2007). A technique for simulating visual field losses in virtual environments to study human navigation. *Behavioral Research Methods*, *39*(3), 552–560. doi:10.3758/BF03193025.

Foulsham, T., Teszka, R., & Kingstone, A. (2011). Saccade control in natural images is shaped by the information visible at fixation: Evidence from asymmetric gaze-contingent windows. *Attention, Perception, & Psychophysics*, *73*, 266–283. doi:10.3758/s13414-010-0014-5.

Foxe, J. J., & Simpson, G. V. (2002). Flow of activation from V1 to frontal cortex in humans. *Experimental Brain Research*, *142*(1), 139–150. doi:10.1007/s00221-001-

0906-7.

Fum, D., Del Missier, F., & Stocco, A. (2007). The cognitive modeling of human behavior: Why a model is (sometimes) better than 10,000 words. *Cognitive Systems Research*, *8*(3), 135–142. doi:10.1016/j.cogsys.2007.07.001.

Gardiner, C. W. (2004). *Handbook of stochastic methods.* Berlin: Springer.

Gattass, R., Gross, C. G., & Sandell, J. H. (1981). Visual topography of V2 in the macaque. *The Journal of Comparative Neurology*, *201*, 519–539. doi:10.1002/cne.902010405.

Gattass, R., Sousa, A. P. B., & Gross, C. G. (1988). Visuotopic organization and extent of V3 and V4 of the macaque. *The Journal of Neuroscience*, *8*, 1831–1845. Retrieved from http://www.jneurosci.org/content/8/6/1831.full.pdf.

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*, 858–873. doi:10.1167/6.9.1.

Gersch, T. M., Kowler, E., & Dosher, B. (2004). Dynamic allocation of visual attention during the execution of sequences of saccades. *Vision Research*, *44*(12), 1469-83. doi:10.1016/j.visres.2003.12.014.

Gilchrist, I. D. (2011). Saccades. In S. P. Liversedge, I. D. Gilchrist, & S. Everling (Eds.), *The Oxford handbook of eye movements* (pp. 85–94). Oxford: Oxford University Press.

Gillespie, D. T. (1976). A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, *22*(4), 403–434. doi:10.1016/0021-9991(76)90041-3.

Glaholt, M. G., Rayner, K., & Reingold, E. M. (2013). Spatial frequency filtering and the direct control of fixation durations during scene viewing. *Attention, Perception, & Psychophysics*, *75*, 1761–1773. doi:10.3758/s13414-013-0522-1.

Goldberg, D. E. (1989). *Genetic algorithms in search, optimization, and machine learning.* Reading, MA: Addison-Wesley.

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, *15*(1), 20–25. doi:10.1016/0166-2236(92)90344-8.

Greene, H. H., Pollatsek, A., Masserang, K., Lee, Y. J., & Rayner, K. (2010). Directional processing within the perceptual span during visual target localization. *Vision Research*, *50*, 1274–1282. doi:10.1016/j.visres.2010.04.012.

Hanes, D. P., & Schall, J. D. (1996). Neural control of voluntary movement initiation. *Science*, *274*(5286), 427–430. doi:10.1126/science.274.5286.427.

Heathcote, A., Brown, S., & Mewhort, D. J. K. (2002). Quantile maximum likelihood estimation of response time distributions. *Psychonomic Bulletin & Review*, *9*(2), 394–401. doi:10.3758/BF03196299.

Henderson, J. M. (1992). Visual attention and eye movement control during reading and picture viewing. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 260–283). New York: Springer.

Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences*, *7*(11), 498–504. doi:10.1016/j.tics.2003.09.006.

Henderson, J. M. (2011). Eye movements and scene perception. In S. P. Liversedge, I. D. Gilchrist, & S. Everling (Eds.), *The Oxford handbook of eye movements* (pp. 593–606). Oxford: Oxford University Press.

Henderson, J. M., Brockmole, J. R., Castelhano, M. S., & Mack, M. (2007). Visual saliency does not account for eye movements during visual search in real-world scenes. In R. P. G. van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 537–562). Oxford: Elsevier.

Henderson, J. M., Chanceaux, M., & Smith, T. J. (2009). The influence of clutter on real-world scene search: Evidence from search efficiency and eye movements. *Journal of Vision*, *9(1):32*, 1–8. doi:10.1167/9.1.32.

Henderson, J. M., & Ferreira, F. (1993). Eye movement control during reading: Fixation measures reflect foveal but not parafoveal processing difficulty. *Canadian Journal of Experimental Psychology*, *47*(2), 201–221. doi:10.1037/h0078814.

Henderson, J. M., & Hollingworth, A. (1998). Eye movements during scene viewing: An overview. In G. D. M. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 269–295). Oxford, UK: Elsevier.

Henderson, J. M., & Hollingworth, A. (1999a). High-level scene perception. *Annual Review of Psychology*, *50*, 243–271. doi:10.1146/annurev.psych.50.1.243.

Henderson, J. M., & Hollingworth, A. (1999b). The role of fixation position in detecting scene changes across saccades. *Psychological Science*, *10*(5), 438–443. doi:10.1111/1467-9280.00183.

Henderson, J. M., Malcolm, G. L., & Schandl, C. (2009). Searching in the dark: Cognitive relevance drives attention in real-world scenes. *Psychonomic Bulletin & Review*, *16*(5), 850–856. doi:10.3758/PBR.16.5.850.

Henderson, J. M., McClure, K. K., Pierce, S., & Schrock, G. (1997). Object identification without foveal vision: Evidence from an artificial scotoma paradigm. *Perception & Psychophysics*, *59*(3), 323–346. doi:10.3758/BF03211901.

Henderson, J. M., Nuthmann, A., & Luke, S. G. (2013). Eye movement control during scene viewing: Immediate effects of scene luminance on fixation durations. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(2), 318–

322. doi:10.1037/a0031224.

Henderson, J. M., Olejarczyk, J., Luke, S. G., & Schmidt, J. (2014). Eye movement control during scene viewing: Immediate degradation and enhancement effects of spatial frequency filtering. *Visual Cognition*, *22*(3-4), 486–502. doi:10.1080/13506285.2014.897662.

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. doi:10.3758/PBR.15.3.566.

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. doi:10.1080/13506280802685552.

Henderson, J. M., Weeks, P. A., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(1), 210–228. doi:10.1037/0096-1523.25.1.210.

Henderson, J. M., Williams, C. C., Castelhano, M. S., & Falk, R. J. (2003). Eye movements and picture processing during recognition. *Perception & Psychophysics*, *65*(5), 725–734. doi:10.3758/BF03194809.

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. doi:10.1038/nn.2669.

Hilz, R., & Cavonius, C. R. (1974). Functional organization of the peripheral retina: Sensitivity to periodic stimuli. *Vision Research*, *14*, 1333–1337. doi:10.1016/0042-6989(74)90006-6.

Hoffman, J. E., & Subramaniam, B. (1995). The role of visual attention in saccadic eye movements. *Perception & Psychophysics*, *57*(6), 787–795. doi:10.3758/BF03206794.

Holland, J. H. (1975). *Adaptation in natural and artificial systems.* Ann Arbor: University of Michigan Press.

Hollingworth, A., & Henderson, J. M. (2002). Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(1), 113–136. doi:10.1037/0096-1523.28.1.113.

Hollingworth, A., Schrock, G., & Henderson, J. M. (2001). Change detection in the flicker paradigm: The role of fixation position within the scene. *Memory & Cognition*, *29*(2), 296–304. doi:10.3758/BF03194923.

Holmes, D. L., Cohen, K. M., Haith, M. M., & Morrison, F. J. (1977). Peripheral visual processing. *Perception & Psychophysics*, *22*(6), 571–577.

doi:10.3758/BF03198765.

Hooge, I. T. C., & Erkelens, C. J. (1996). Control of fixation duration in a simple search task. *Perception & Psychophysics*, *58*(7), 969–976. doi:10.3758/BF03206825.

Hooge, I. T. C., & Erkelens, C. J. (1998). Adjustment of fixation duration in visual search. *Vision Research*, *38*(9), 1295–1302. doi:10.1016/S0042-6989(97)00287-3.

Hooge, I. T. C., & Erkelens, C. J. (1999). Peripheral vision and oculomotor control during visual search. *Vision Research*, *39*(8), 1567–1575. doi:10.1016/S0042-6989(98)00213-2.

Ho-Phuoc, T., Guyader, N., Landragin, F., & Guérin-Dugué, A. (2012). When viewing natural scenes, do abnormal colors impact on spatial or temporal parameters of eye movements? *Journal of Vision*, *12(2):4*, 1–13. doi:10.1167/12.2.4.

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. doi:10.1113/jphysiol.1959.sp006308.

Hwang, A. D., Higgins, E. C., & Pomplun, M. (2009). A model of top-down attentional control during visual search in complex scenes. *Journal of Vision*, *9(5):25*, 1–18. doi:10.1167/9.5.25.

Hwang, A. D., Wang, H.-C., & Pomplun, M. (2011). Semantic guidance of eye movements in real-world scenes. *Vision Research*, *51*(10), 1192–1205. doi:10.1016/j.visres.2011.03.010.

Ikeda, M., & Takeuchi, T. (1975). Influence of foveal load on the functional visual field. *Perception & Psychophysics*, *18*(4), 255–260. doi:10.3758/BF03199371.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*(10-12), 1489–1506. doi:10.1016/S0042-6989(99)00163-7.

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20*(11), 1254–1259. doi:10.1109/34.730558.

Jacobs, A. M. (1986). Eye-movement control in visual search: How direct is visual span control? *Perception & Psychophysics*, *39*(1), 47–58. doi:10.3758/BF03207583.

Jones, L. A., & Higgins, G. C. (1947). Photographic granularity and graininess. III. Some characteristics of the visual system of importance in the evaluation of graininess and granularity. *Journal of the Optical Society of America*, *37*(4), 217–263. doi:10.1364/JOSA.37.000217.

Juola, J. F., Bouwhuis, D. G., Cooper, E. E., & Warner, C. B. (1991). Control of attention around the fovea. *Journal of Experimental Psychology: Human Perception and Performance*, *17*(1), 125–141. doi:10.1037/0096-1523.17.1.125.

Kienzle, W., Franz, M. O., Schölkopf, B., & Wichmann, F. A. (2009). Center-surround patterns emerge as optimal predictors for human saccade targets. *Journal of Vision*, *9(5):7*, 1–15. doi:10.1167/9.5.7.

King-Smith, P. E., Grigsby, S. S., Vingrys, A. J., Benes, S. C., & Supowit, A. (1994). Efficient and unbiased modifications of the QUEST threshold method: Theory, simulations, experimental evaluation and practical implementation. *Vision Research*, *34*(7), 885–912. doi:10.1016/0042-6989(94)90039-6.

Kleiner, M., Brainard, D. H., & Pelli, D. G. (2007). What's new in Psychtoolbox-3? *Perception*, *36*, 14.

Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying circuitry. *Human Neurobiology*, *4*, 219–227. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/3836989.

Kowler, E., Anderson, E., Dosher, B., & Blaser, E. (1995). The role of attention in the programming of saccades. *Vision Research*, *35*(13), 1897–1916. doi:10.1016/0042-6989(94)00279-U.

LaBerge, D., & Brown, V. (1989). Theory of attentional operations in shape identification. *Psychological Review*, *96*(1), 101–124. doi:10.1037/0033-295X.96.1.101.

Land, M. F. (1999). Motion and vision: why animals move their eyes. *Journal of Comparative Physiology A*, *185*(4), 341–352. doi:10.1007/s003590050393.

Larson, A. M., & Loschky, L. C. (2009). The contributions of central versus peripheral vision to scene gist recognition. *Journal of Vision*, *9(10):6*, 1–16. doi:10.1167/9.10.6.

Laubrock, J., Cajar, A., & Engbert, R. (2013). Control of fixation duration during scene viewing by interaction of foveal and peripheral processing. *Journal of Vision*, *13(12):11*, 1–20. doi:10.1167/13.12.11.

Lewandowsky, S., & Farrell, S. (2011). *Computational modeling in cognition: Principles and practice.* Los Angeles [u.a.]: Sage.

Lingnau, A., Schwarzbach, J., & Vorberg, D. (2010). (Un-)Coupling gaze and attention outside central vision. *Journal of Vision*, *10(11):13*, 1–13. doi:10.1167/10.11.13.

Loftus, G. R. (1985). Picture perception: Effects of luminance on available information and information-extraction rate. *Journal of Experimental Psychology: General*, *114*(3), 342–356. doi:10.1037/0096-3445.114.3.342.

Loftus, G. R., Kaufman, L., Nishimoto, T., & Ruthruff, E. (1992). Effects of visual degradation on eye-fixation duration, perceptual processing, and long-term visual memory. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 203–226). New York: Springer.

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. doi:10.1037/0096-1523.4.4.565.

Loschky, L. C., & McConkie, G. W. (2002). Investigating spatial vision and dynamic attentional selection using a gaze-contingent multiresolutional display. *Journal of Experimental Psychology: Applied*, *8*(2), 99–117. doi:10.1037//1076-898X.8.2.99.

Loschky, L. C., McConkie, G. W., Yang, J., & Miller, M. E. (2005). The limits of visual resolution in natural scene viewing. *Visual Cognition*, *12*(6), 1057–1092. doi:10.1080/13506280444000652.

Ludwig, C. J. H., Davies, J. R., & Eckstein, M. P. (2014). Foveal analysis and peripheral selection during active visual sampling. *Proceedings of the National Academy of Sciences of the United States of America*, *111*, E291–E299. doi:10.1073/pnas.1313553111.

Luke, S. G., Nuthmann, A., & Henderson, J. M. (2013). Eye movement control in scene viewing and reading: Evidence from the stimulus onset delay paradigm. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(1), 10–15. doi:10.1037/a0030392.

Mackworth, N. H. (1965). Visual noise causes tunnel vision. *Psychonomic Science*, *3*(1-12), 67–68. doi:10.3758/BF03343023.

Malcolm, G. L., & Shomstein, S. (2015). Object-based attention in real-world scenes. *Journal of Experimental Psychology: General*, *144*(2), 257–263. doi:10.1037/xge0000060.

Mannan, S. K., Ruddock, K. H., & Wooding, D. S. (1995). Automatic control of saccadic eye movements made in visual inspection of briefly presented 2-D images. *Spatial Vision*, *9*(3), 363–386. doi:10.1163/156856895X00052.

Mannan, S. K., Ruddock, K. H., & Wooding, D. S. (1996). The relationship between locations of spatial features and those of fixations made during visual examination of briefly presented images. *Spatial Vision*, *10*(3), 165–188. doi:10.1163/156856896X00123.

Mannan, S. K., Ruddock, K. H., & Wooding, D. S. (1997). Fixation patterns made during brief examination of two-dimensional images. *Perception*, *26*(8), 1059–1072. doi:10.1068/p261059.

Matin, E. (1974). Saccadic suppression: A review and an analysis. *Psychological Bulletin*, *81*(12), 899–917. doi:10.1037/h0037368.

McConkie, G. W., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception & Psychophysics*, *17*(6), 578–586. doi:10.3758/BF03203972.

Mills, M., Hollingworth, A., Van der Stigchel, S., & Hoffman, M. D., L. Dodd. (2011).

Examining the influence of task set on eye movements and fixations. *Journal of Vision*, *11(8):17*, 1–15. doi:10.1167/11.8.17.

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorial in Quantitative Methods for Psychology*, *42*(2), 61–64. Retrieved from http://pcl.missouri.edu/sites/default/files/morey.2008.pdf.

Morrison, R. E. (1984). Manipulation of stimulus onset delay in reading: Evidence for parallel programming of saccades. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(5), 667–682. doi:10.1037/0096-1523.10.5.667.

Müller, N. G., Bartelt, O. A., Donner, T. H., Villringer, A., & Brandt, S. A. (2003). A physiological correlate of the "zoom lens" of visual attention. *Journal of Neuroscience*, *23*(9), 3561–3565. Retrieved from http://www.jneurosci.org/content/23/9/3561.full.pdf+html.

Nakayama, K., & MacKeben, M. (1989). Sustained and transient components of focal visual attention. *Vision Research*, *29*(11), 1631–1647. doi:10.1016/0042-6989(89)90144-2.

Navalpakkam, V., & Itti, L. (2005). Modeling the influence of task on attention. *Vision Research*, *45*, 205–231. doi:10.1016/j.visres.2004.07.042.

Neider, M. B., & Zelinsky, G. J. (2006). Scene context guides eye movements during visual search. *Vision Research*, *10*(5), 614–621. doi:10.1016/j.visres.2005.08.025.

Nelson, W. W., & Loftus, G. R. (1980). The functional visual field during picture viewing. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(4), 391–399. doi:10.1037/0278-7393.6.4.391.

Nuthmann, A. (2013). On the visual span during object search in real-world scenes. *Visual Cognition*, *21*(7), 803–837. doi:10.1080/13506285.2013.832449.

Nuthmann, A. (2014). How do the regions of the visual field contribute to object search in real-world scenes? Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(1), 342–360. doi:10.1037/a0033854.

Nuthmann, A. (2016). Fixation durations in scene viewing: Modeling the effects of local image features, oculomotor parameters, and task. *Psychonomic Bulletin & Review*, 1–23. Advance online publication. doi:10.3758/s13423-016-1124-4.

Nuthmann, A., & Einhäuser, W. (2015). A new approach to modeling the influence of image features on fixation selection in scenes. *Annals of the New York Academy of Sciences*, *1339*, 82–96. doi:10.1111/nyas.12705.

Nuthmann, A., & Engbert, R. (2009). Mindless reading revisited: An analysis based on the SWIFT model of eye-movement control. *Vision Research*, *49*(3), 322–336. doi:10.1016/j.visres.2008.10.022.

Nuthmann, A., & Malcolm, G. L. (2016). Eye-guidance during real-world scene search: The role color plays in central and peripheral vision. *Journal of Vision*, *16(2):3*, 1–16. doi:10.1167/16.2.3.

Nuthmann, A., Smith, T. J., Engbert, R., & Henderson, J. M. (2010). CRISP: A computational model of fixation durations in scene viewing. *Psychological Review*, *117*(2), 382–405. doi:10.1037/a0018924.

Pannasch, S., Schulz, J., & Velichovsky, B. M. (2011). On the control of visual fixation durations in free viewing of complex images. *Attention, Perception, & Psychophysics*, *73*(4), 1120–1132. doi: 10.3758/s13414-011-0090-1.

Parkhurst, D., Culurciello, E., & Niebur, E. (2000). Evaluating variable resolution displays with visual search: Task performance and eye movements. In A. T. Duchowski (Ed.), *Proceedings of the eye tracking research and applications symposium 2000* (pp. 105–109. doi:10.1145/355017.355033). New York: Association of Computing Machinery.

Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual selective attention. *Vision Research*, *42*(1), 107–123. doi:10.1016/S0042-6989(01)00250-4.

Pelli, D. G. (1997). The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442. doi:10.1163/156856897X00366.

Pomplun, M., Reingold, E. M., & Shen, J. (2001). Peripheral and parafoveal cueing and masking effects on saccadic selectivity in a gaze-contingent window paradigm. *Vision Research*, *41*(21), 2757–2769. doi:10.1016/S0042-6989(01)00145-6.

Posner, M. I., Snyder, C. R. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*, *109*(2), 160–174. doi:10.1037/0096-3445.109.2.160.

R Core Team. (2015). *R: A language and environment for statistical computing.* Vienna: R Foundation for Statistical Computing. URL: http://www.R-project.org/.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422. doi:10.1037/0033-2909.124.3.372.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, *62*(8), 1457–1506. doi:10.1080/17470210902816461.

Rayner, K., & Bertera, J. H. (1979). Reading without a fovea. *Science*, *206*, 468–469. doi:10.1126/science.504987.

Rayner, K., & McConkie, G. W. (1976). What guides a reader's eye movements?

Vision Research, *16*(8), 829–837. doi:10.1016/0042-6989(76)90143-7.

Rayner, K., & Pollatsek, A. (1981). Eye movement control during reading: Evidence for direct control. *Quarterly Journal of Experimental Psychology*, *33A*(4), 351–373. doi:10.1080/14640748108400798.

Rayner, K., Slowiaczek, M. L., Clifton, C., & Bertera, J. H. (1983). Latency of sequential eye movements: Implications for reading. *Journal of Experimental Psychology: Human Perception and Performance*, *9*(6), 912–922. doi:10.1037/0096-1523.9.6.912.

Reddi, B. A. J., & Carpenter, R. H. S. (2000). The influence of urgency on decision time. *Nature Neuroscience*, *3*(8), 827–830. doi:10.1038/77739.

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, *105*(1), 125–157. doi:10.1037/0033-295X.105.1.125.

Reinagel, P., & Zador, A. M. (1999). Natural scene statistics at the center of gaze. *Network: Computation in Neural Systems*, *10*, 341–350. doi:10.1088/0954-898x/10/4/304.

Reingold, E. M., & Loschky, L. C. (2002). Saliency of peripheral targets in gaze-contingent multiresolutional displays. *Behavioral Research Methods, Instruments, & Computers*, *34*(4), 491–499. doi:10.3758/BF03195478.

Reingold, E. M., Loschky, L. C., McConkie, G. W., & Stampe, D. M. (2003). Gaze-contingent multiresolutional displays: An integrative review. *Human Factors*, *45*(2), 307–328. doi:10.1518/hfes.45.2.307.27235.

Reingold, E. M., Reichle, E. D., Glaholt, M. G., & Sheridan, H. (2012). Direct lexical control of eye movements in reading: Evidence from a survival analysis of fixation durations. *Cognitive Psychology*, *65*, 177–206. doi:10.1016/j.cogpsych.2012.03.001.

Ringer, R. V., Throneburg, Z., Johnson, A. P., Kramer, A. F., & Loschky, L. C. (2016). Impairing the useful field of view in natural scenes: Tunnel vision versus general interference. *Journal of Vision, Special Issue: Scene perception from central to peripheral vision*, *16(2):7*, 1–25, doi:10.1167/16.2.7.

Robson, J. G., & Graham, N. (1981). Probability summation and regional variation in contrast sensitivity across the visual field. *Vision Research*, *21*(3), 409–418. doi:10.1016/0042-6989(81)90169-3.

Rovamo, J., & Virsu, V. (1979). An estimation and application of the human cortical magnification factor. *Experimental Brain Research*, *37*(3), 495–510. doi:10.1007/BF00236819.

Rovamo, J., Virsu, V., & Näsänen, R. (1978). Cortical magnification factor predicts the photopic contrast sensitivity of peripheral vision. *Nature*, *271*, 54–56. doi:10.1038/271054a0.

Sagi, D., & Julesz, B. (1985). Fast noninertial shifts of attention. *Spatial Vision*, *1*(2), 141–149. doi:10.1163/156856885X00152.

Saida, S., & Ikeda, M. (1979). Useful visual field size for pattern perception. *Perception & Psychophysics*, *25*(2), 119–125. doi:10.3758/BF03198797.

Shapley, R., & Lennie, P. (1985). Spatial frequency analysis in the visual system. *Annual Review of Neuroscience*, *8*, 547–583. doi:10.1146/annurev.ne.08.030185.002555.

Shioiri, S. (1993). Postsaccadic processing of the retinal image during picture scanning. *Perception & Psychophysics*, *53*(3), 305–314. doi:10.3758/BF03205185.

Shioiri, S., & Ikeda, M. (1989). Useful resolution for picture perception as a function of eccentricity. *Perception*, *18*(3), 347–361. doi:10.1068/p180347.

Silverman, B. W. (1986). *Density estimation.* London: Chapman and Hall.

Smith, P. L., & Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends in Neurosciences*, *27*(3), 161–168. doi:10.1016/j.tins.2004.01.006.

Smith, T. J., & Henderson, J. M. (2009). Facilitation of return during scene viewing. *Visual Cognition*, *17*(6-7), 1083–1108. doi:10.1080/13506280802678557.

Smith, T. J., & Henderson, J. M. (2011). Does oculomotor inhibition of return influence fixation probability during scene search? *Attention, Perception, & Psychophysics*, *73*, 2384–2398. doi:10.3758/s13414-011-0191-x.

Spotorno, S., Malcolm, G. L., & Tatler, B. W. (2014). How context information and target information guide the eyes from the first epoch of search in real-world scenes. *Journal of Vision*, *14(2):7*, 1–21. doi:10.1167/14.2.7.

Strasburger, H., Rentschler, I., & Jüttner, M. (2011). Peripheral vision and pattern recognition: A review. *Journal of Vision*, *11(5):13*, 1–82. doi:10.1167/11.5.13.

Tatler, B. W. (2007). The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision*, *7(14):4*, 1–17. doi:10.1167/7.14.4.

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):5*, 1–23. doi:10.1167/11.5.5.

Tatler, B. W., & Vincent, B. T. (2008). Systematic tendencies in scene viewing. *Journal of Eye Movement Research*, *2*(2), 1–18. doi:10.16910/jemr.2.2.5.

Tatler, B. W., & Vincent, B. T. (2009). The prominence of behavioural biases in eye guidance. *Visual Cognition*, *17*(6-7), 1029–1054. doi:10.1080/13506280902764539.

Thorpe, S. J., Gegenfurtner, K. R., Fabre-Thorpe, M., & Bülthoff, H. H. (2001).

Detection of animals in natural images using far peripheral vision. *European Journal of Neuroscience*, *14*(5), 869–876. doi:10.1046/j.0953-816x.2001.01717.x.

Tootell, R. B., Silverman, M. S., & De Valois, R. L. (1981). Spatial frequency columns in primary visual cortex. *Science*, *214*(4522), 813–815. doi:10.1126/science.7292014.

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. doi:10.1037/0033-295X.113.4.766.

Trukenbrod, H. A., & Engbert, R. (2012). Eye movements in a sequential scanning task: Evidence for distributed processing. *Journal of Vision*, *12(1):5*, 1–12. doi:10.1167/12.1.5.

Trukenbrod, H. A., & Engbert, R. (2014). ICAT: A computational model for the adaptive control of fixation durations. *Psychonomic Bulletin & Review*, *21*(4), 907–934. doi:10.3758/s13423-013-0575-0.

Tsotsos, J. K., Culhane, S. M., Wai, W. Y. K., Lai, Y., Davis, N., & Nuflo, F. (1995). Modeling visual attention via selective tuning. *Artificial Intelligence*, *78*(1-2), 507–545. doi:10.1016/0004-3702(95)00025-9.

Underwood, G., & Foulsham, T. (2006). Visual saliency and semantic incongruency influence eye movements when inspecting pictures. *Quarterly Journal of Experimental Psychology*, *59*(11), 1931–1949. doi:10.1080/17470210500416342.

van Diepen, P. M. J. (2001). Foveal stimulus degradation during scene perception. In F. H. Columbus (Ed.), *Advances in psychology research* (Vol. 2, pp. 89–115). Huntington, NY: Nova Science Publishers.

van Diepen, P. M. J., De Graef, P., & d'Ydewalle, G. (1995). Chronometry of foveal information extraction during scene perception. In J. M. Findlay, R. Walker, & R. W. Kentridge (Eds.), *Eye movement research: Mechanisms, processes and applications* (pp. 349–362). NewYork, NY: Elsevier.

van Diepen, P. M. J., & d'Ydewalle, G. (2003). Early peripheral and foveal processing in fixations during scene perception. *Visual Cognition*, *27*(1), 1141–1151. doi:10.1080/713756668.

van Diepen, P. M. J., & Wampers, M. (1998). Scene exploration with Fourier-filtered peripheral information. *Perception*, *27*(10), 1141–1151. doi:10.1068/p271141.

van Kampen, N. G. (1981). *Stochastic processes in physics and chemistry.* Amsterdam: North-Holland.

Velisavljević, L., & Elder, J. H. (2008). Visual short-term memory for natural scenes: Effects of eccentricity. *Journal of Vision*, *8(4):28*, 1–17. doi:10.1167/8.4.28.

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S.* New York:

Springer.

Virsu, V. (1979). Visual resolution, contrast sensitivity, and the cortical magnification factor. *Experimental Brain Research*, *37*(3), 475–494. doi:10.1007/BF00236818.

Vitu, F., Lancelin, D., & Marrier d'Unienville, V. (2007). A perceptual-economy account for the inverted-optimal viewing position effect. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(5), 1220–1249. doi:10.1037/0096-1523.33.5.1220.

Võ, M. L.-H., & Henderson, J. M. (2009). Does gravity matter? Effects of semantic and syntactic inconstistencies on the allocation of attention during scene perception. *Journal of Vision*, *9(3):24*, 1–15. doi:10.1167/9.3.24.

Walshe, R. C., & Nuthmann, A. (2014). Asymmetrical control of fixation durations in scene viewing. *Vision Research*, *100*, 38–46. doi:10.1016/j.visres.2014.03.012.

Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & Psychophysics*, *33*(2), 113–120. doi:10.3758/BF03202828.

Wertheim, T. (1894). Über die indirekte Sehschärfe. *Zeitschrift für Psychologie und Physiologie der Sinnesorgane*, *7*, 121–187.

Williams, L. J. (1985). Tunnel vision induced by a foveal load manipulation. *Human Factors*, *27*(2), 221–227. doi:10.1177/001872088502700209.

Williams, L. J. (1988). Tunnel vision or general interference? cognitive load and attentional bias are both important. *The American Journal of Psychology*, *101*(2), 171–191. doi:10.2307/1422833.

Williams, L. J. (1989). Foveal load affects the functional field of view. *Human Performance*, *2*(1), 1–28. doi:10.1207/s15327043hup0201_1.

Williams, L. J. (1995). Peripheral target recognition and visual field narrowing in aviators and nonaviators. *International Journal of Aviation Psychology*, *5*(2), 215–232. doi:10.1207/s15327108ijap0502_6.

Wilming, N., Harst, S., Schmidt, N., & König, P. (2013). Saccadic momentum and facilitation of return saccades contribute to an optimal foraging strategy. *PLOS Computational Biology*, *9*(1), 1–13. doi:10.1371/journal.pcbi.1002871.

Wischnewski, M., Belardinelli, A., Schneider, W. X., & Steil, J. J. (2010). Where to look next? Combining static and dynamic proto-objects in a TVA-based model of visual attention. *Cognitive Computation*, *2*, 326–343. doi:10.1007/s12559-010-9080-1.

Yang, S.-N., & McConkie, G. W. (2001). Eye movements during reading: a theory of saccade initiation times. *Vision Research*, *41*(25-26), 3567–3585. doi:10.1016/S0042-6989(01)00025-6.

Yarbus, A. L. (1967). *Eye movements and vision.* New York: Plenum Press.

Zelinsky, G. J. (2008). A theory of eye movements during target acquisition. *Psychological Review*, *115*(4), 787–835. doi:10.1037/a0013118.