EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

# Predictive coding and spike timing dependent plasticity in primary visual cortex

**Diploma thesis**

Eberhard Karls Universität
Wilhelm-Schickard-Institut für Informatik
Tübingen

Max-Planck-Institut für biologische Kybernetik
Tübingen

Baylor College of Medicine
Department of Neuroscience
Houston, TX, USA

Vorgelegt von: **Alexander Ecker**

Erstkorrektor: **Prof. Dr. Andreas Schilling**
Betreuer/Zweitkorrektor: **Andreas Tolias, Ph.D.**

Eingereicht am: **24. März 2008**

# Contents

# Abstract

The aim of the present thesis is to elucidate whether the brain uses the statistical regularities of its environment to facilitate faster information processing and to investigate the mechanisms it uses to adapt to changes in these regularities. To this end, psychophysical studies in human and non-human primate are combined with electrophysiological single cell recordings from primary visual cortex (V1) of the macaque monkey.

In the first part, we show that the flash-lag effect—a visual illusion in which the position of a predictably moving object is misjudged relative to that of a briefly flashed object—can also be observed in monkeys. By studying the responses of single neurons to moving and flashed bars, a potential mechanism put forward to explain the flash-lag effect, the differential latencies hypothesis, is rejected.

In the second part, a potential mechanism underlying the dynamic reorganization of neural circuits in response to changing statistical regularities of their environment is examined. This mechanism, spike timing dependent plasticity, has been shown to induce "predictive shifts" of neural population responses after repeated presentation of the same stimulus sequences. We demonstrate that motion conditioning can induce such predictive shifts in the visual system of awake monkeys. In contrast to previous findings, however, they are specific to the exact nature of the stimulus (moving object) rather than being manifested in general receptive field shifts. This could be a sign of a principled mechanism by which circuits in early visual areas can undergo changes as a result of learning without at the same time getting disturbed under conditions that have already been learned in the past.

# 1 Introduction

> "Because neural processing takes time, perceptual systems must draw conclusions about the outside world based on data that are slightly outdated—in other words, perception lives slightly in the past." (Eagleman and Sejnowski 2007)

With this statement, Eagleman and Sejnowski (2007) open their recent article about the flash-lag effect—a fascinating visual illusion in which a moving object's position relative to a briefly flashed one is systematically misjudged by observers. This statement nicely illustrates a key problem in perception, already brought up by Helmholtz more than a century ago (Helmholtz 1896): Is a moving object's perceived position—because of neural processing delays—lagging behind its real position? Or is the visual system somehow compensating its own processing delays, making us see objects where they actually are?

From an evolutionary point of view, such compensation would certainly be very useful, since predicting the trajectory of a moving object is only one of many examples where exploiting statistical regularities of the world could reduce neural processing time. For instance when a monkey is jumping from one tree to another, the visual system only has a few tens of milliseconds to locate the branch to grasp. Similarly, if a prey is trying to escape its predator, a few milliseconds can obviously save its live. On the other hand, if the brain is using statistical regularities of the world to compensate for processing delays or even predict future events, violation of these regularities should result in wrong perceptions—as can be observed in visual illusions.

In this thesis, we seek to address the question whether and how the primate visual system is exploiting statistical regularities of the world in order to speed up processing. In the first part, we study a well-known visual illusion—the above mentioned flash-lag effect. Although various hypothesis about its underlying mechanisms have been put forth by psychophysicists, none of these can account for all aspects of the phenomenon. We therefore combine psychophysical and electrophysiological methods to gain better insights into potential mechanisms generating or contributing to the flash-lag illusion in primary visual cortex. Based on the analysis of single cell responses to moving and flashed stimuli, we provide evidence that one of the proposed mechanisms that predicted stimuli are processed faster than unpredictable ones is not implemented in primary visual cortex.

In the second part we study how the brain adapts to changes in the environment and how it learns new regularities. The recently discovered spike timing dependent plasticity rule (Bi and Poo 1998; Zhang et al. 1998) has been shown theoretically to implement a form of sequence learning in an unsupervised fashion leading to a predictive population response (Abbott and Blum 1996). We will use a simple stimulus with a bar moving across the visual field and parametrically vary the probability of leftwards versus rightwards direction of motion to study how this simple regularity affects responses of V1 neurons. Contrary to previous findings (Yao and Dan 2001; Fu et al. 2002), we find that

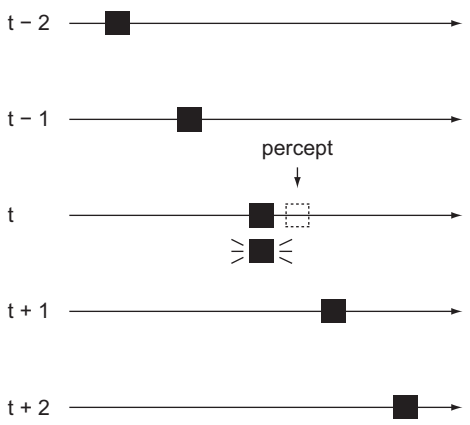


Figure 1.1: Flash-lag illusion. The top square is moving at constant speed from left to right. At time $t$, a second square is flashed below at exactly the same horizontal location for just one monitor frame. Thus, for the entire time the flashed square is visible, the moving square is at the same physical location. Yet do observes report the moving square to be ahead of the flashed one as depicted by the dashed square.

predictive shifts in the population response are specific to the stimulus nature, i. e. does not transfer from moving to flashed bars.

## 1.1 The flash-lag effect

In the flash-lag illusion, subjects compare the position of a moving object to that of a briefly flashed one. While the first object (typically a bar or a dot) is moving across the visual field at constant speed, the second object is flashed below the first one for a short time. Although, physically, the two objects are perfectly aligned, observers report that they perceive the moving object ahead of the flashed one. Fig. 1.1 illustrates the flash-lag stimulus and the perceptual illusion.

The flash-lag illusion has first been reported by Hazelhoff and Wiersma in 1924. After it had been rediscovered by Nijhawan (1994), it celebrated a revival in more recent years (for reviews, see Krekelberg and Lappe 2001; Nijhawan 2002) and created a substantial debate about its underlying mechanisms (Krekelberg et al. 2000; Patel et al. 2000). In this debate, various modifications to the original flash-lag experiment described above (sometimes called the *continuous motion condition*) played an important role. Among those, the *flash-initiated condition* (FIC), the *flash-terminated condition* (FTC), and the *flash reversal condition* (FRC) are the most prominent ones. In the flash-initiated condition, moving and flashed object appear at the same point in time and while one object starts moving, the other immediately disappears again. In the flash-terminated condition, the flash occurs in the last frame of the motion and both objects disappear at the same point in time. Finally, in the flash reversal condition, the moving bar reverses its direction at the time of the flash, or slightly before or after it.

In the ongoing debate about the origin of the flash-lag effect, three main hypotheses have been put forward, which we will briefly discuss below.

### 1.1.1 The extrapolation hypothesis

Because processing of sensory stimuli inevitably involves neural delays, the internal representation of the world in the neural circuits of the brain can never be fully up to date. In some cases, however, it is possible to compensate for the delays by exploiting the statistics of the natural environment: As motion is generally smooth, a moving object will most likely continue to move from one time instant to the next. In the *extrapolation* view, neural circuits actively compensate for delays by extrapolating the past trajectory of a moving object (Nijhawan 1994). For a flashed object, naturally no such compensation can occur. This concept is illustrated in Fig. 1.2 A: since a prediction is extracted from the neural signal about the location of the moving stimulus, but no such prediction can be made for the flashed object, at the time of the flash the moving object is perceived leading the flashed one.

In this view, the flash-lag effect has been hypothesized to involve neurons in early visual areas whose receptive fields might undergo a "predictive remapping": receptive field locations shift towards the direction from which a moving bar came (Nijhawan 1997; **?**). There is, however, strong evidence against this simple version of the prediction theory: Whitney and Murakami (1998) and Whitney et al. (2000) showed that when the moving bar reverses its direction of motion, the location of the flash is perceived to be aligned with the moving bar, which is inconsistent with the extrapolation theory. Generally, the perceived location of the moving bar is not seen to overshoot its reversal point, as would be expected in the extrapolation model (but see Kanai et al. 2004).

### 1.1.2 The differential latencies hypothesis

An alternative account of the flash-lag effect holds that moving and flashed objects are processed at *different latencies*. Because cells at one location facilitate the responses of other detectors in the vicinity of the moving object, they require less input to be excited. Specifically, it has been put forward to explain the results of Whitney and Murakami (1998) and Whitney et al. (2000): Under this model, the perceived trajectory of the moving bar is identical to the true trajectory but shifted in time. As the moving bar is subject to smaller response latencies, the shift of the trajectory of the moving bar is smaller than that of the flashed bar, as depicted in Fig. 1.2 B. In contrast to the extrapolation model, the perceived location of the moving bar is never predicted to overshoot its real location in the FTC and FRC. Often differential latencies are combined with a temporal averaging procedure to account for the observed smoothing of the curves in (Whitney and Murakami 1998; Whitney et al. 2000).

This account of the flash-lag effect has been criticized for several reasons. First, in the flash-initiation condition, the flash-lag effect is reported to be of the same magnitude as under standard conditions (Eagleman and Sejnowski 2000). However, both objects appear at the same point in time and motion only starts afterwards, ruling out different latencies. Second, under some circumstances, the moving object is perceived at locations it never actually occupied. For instance, Kanai et al. (2004) report perceptual overshoot of the moving object in the FTC under uncertain conditions. Additionally, Eagleman and Sejnowski (2007) show that a flash-lag effect can also be created with apparent motion (see below). Since the differential latencies model postulates that objects are always perceived at their true spatial location but shifted in time, it cannot account for those mislocalizations.

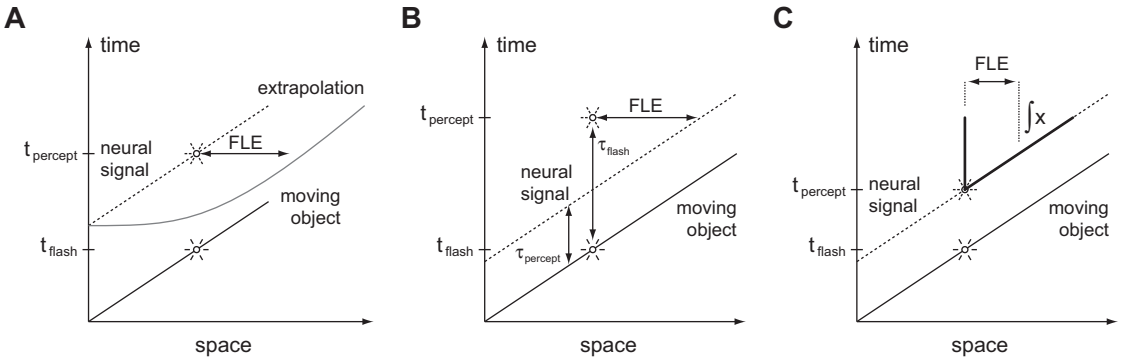


Figure 1.2: Schematic illustration of the different explanations for the flash lag effect (FLE). Space/time diagrams of a moving and a flashed object are shown. The moving object is represented by a black line, the flash by an asterisk. The neural signal is depicted by a dashed line. **A:** *Extrapolation model.* Flashed and moving object are processed with equal neural delay, but the position of the moving object is extrapolated, resulting in the flash-lag effect. Note the spatial overshoot in the flash-termination condition predicted by this model. **B:** *Differential latencies model.* Moving and flashed object are processed with different neural delays. Thus, by the time the flash is perceived, the internal representation of the moving object is already ahead. **C:** *Motion integration model.* Upon a flash, the internal model of the world (moving object) is devalued and a new model acquired. Since the object locations contributing to the new model are all ahead of the flash location, the perceived location of the moving object is biased.

### 1.1.3 Motion integration and position bias

The *temporal averaging model* proposes that the flash-lag illusion is a result of the visual system integrating position information over small time windows. The position of the two objects at a given time is not computed instantaneously, but taken to be their mean position in the integration frame (Lappe and Krekelberg 1998). For the flashed bar, this is its true position, but for the moving bar this is an average over all positions it occupied in the time window (see Fig. 1.2 C). The length of the time window has been reported to range from 50 to 150 ms—in some cases even 500 ms (Krekelberg and Lappe 2001).

This model explains well why position estimates of the moving bar are influenced by events that happened after the flash has occurred. Eagleman and Sejnowski (2007) have demonstrated this effect quite convincingly. Observers were to judge the position of a flash relative to a moving dot. The perception of movement, however, was created in two different ways: In one condition, the dot was moved via five stations to its final location; in the other condition, it jumped from its initial location to the final one. Both conditions created a perceived flash-lag, whereas in the second condition, the dot was perceived at a location it never actually occupied. This argues strongly against a differential latency model, since under this model perceived locations correspond only to true locations but shifted in time. Also, the movement information is not available until long after the flash has occurred, as only then the apparent motion created by the jumping dot can be perceived.

A variant of this scheme is known as *postdiction* (Eagleman and Sejnowski 2000) or *position bias model* (Eagleman and Sejnowski 2007). To construct a percept, the brain constantly combines external input with an internal model of the world, which is based

on the recent past. Under normal circumstances, the internal model is heavily relied upon—as in the prediction account. However, if something unexpected such as a flash occurs, the internal model is devalued and, based on incoming input, an updated internal model is constructed. Thus, the new model is biased towards positions on the trajectory following the flash and the flash appears to be lagging behind the movement. (Eagleman and Sejnowski 2007).

However, this model fails to explain some effects of luminance dependence of the flash-lag effect (Purushothaman et al. 1998; Patel et al. 2000). If the flashed object is presented at much higher luminance than the moving object, the flash-lag effect turns into a flash-lead effect. This cannot be explained by the motion integration model since even if the internal model is not devalued at all, the moving bar is never predicted to trail the flashed bar.

### 1.1.4 Related work

In humans, the flash-lag effect is psychophysically well studied. A large amount of data is available describing various parameters influencing the presence, absence, magnitude, or direction of the effect under different conditions (Nijhawan 1994, 1997; Krekelberg and Lappe 1999; Lappe and Krekelberg 1998; Kanai et al. 2004; Sheth et al. 2000; Watanabe et al. 2001, 2002, 2003; Whitney and Murakami 1998; Whitney et al. 2000; Purushothaman et al. 1998; Eagleman and Sejnowski 2000, 2007; Patel et al. 2000). In contrast, up to now neurophysiological evidence in favor of one model or the other is rather sparse. This is partly owed to the fact that—with the exception of the differential latencies hypothesis—it is not obvious how to neurophysiologically verify or reject these hypotheses. The differential latencies hypothesis, however, makes a clear prediction about the behavior of single neurons and can, therefore, be tested experimentally. In fact, even though it does not appear to be able to account for all psychophysical findings, there is evidence for faster processing of moving stimuli compared to flashed ones in early visual areas. Berry et al. (1999) report anticipatory firing of frog and salamander retinal ganglion cells in response to moving stimuli. Jancke et al. (2004) find shorter latencies for moving stimuli than for flashes in cat primary visual cortex.

## 1.2 Spike timing-dependent plasticity

In the first part a rather universal statistical regularity is considered—namely that motion is usually smooth and objects do not suddenly disappear unless they become occluded by other objects. Since this is such a universal rule, it is probably built into the neural circuit by evolution or early in development. In the second part of this thesis we will turn to the question of how neural circuits could *learn* the statistical regularities of the world in order to reduce processing delays or generate predictions about future events.

By learning we here mean experience-dependent modification of synapses in neural circuits. More than half a century ago, Hebb (1949) suggested a simple yet powerful rule for such synaptic modifications: He postulated that if two neurons fire action potentials in close temporal proximity, synapses between those neurons are strengthened. Since in this type of rule, synaptic modification depends on the temporal correlation of neural activity, it is often referred to as a correlational learning rule. Subsequently, the dis-

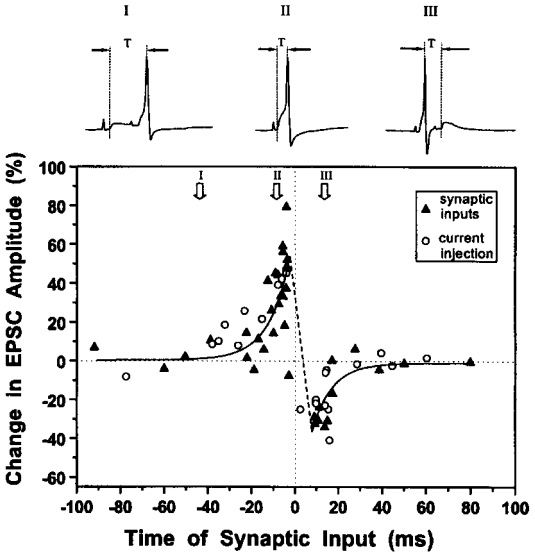


Figure 1.3: Synaptic modification as a function of temporal difference between pre- and postsynaptic action potentials. The $x$-axis depicts the time of arrival of synaptic input relative to the postsynaptic action potential. If the presynaptic event occurs slightly before the postsynaptic action potential, synaptic potentiation is strongest. If the order is reversed, however, the synapse is weakened. Figure from Zhang et al. (1998)

covery of long-term potentiation (LTP; Bliss and Lomo 1973) lent experimental support to Hebb's postulate. However, in its original form, the Hebbian learning rule would not be stable, as synaptic strengths would increase without bound but never decrease. Additionally, the fact that under certain stimulation conditions, long-term depression of synapses can be induced (Ito 1986, 1989) is in disagreement with Hebb's original formulation. Although Stent (1973) suggested a decrease of synaptic strength when correlated firing is absent, more recently it has been discovered that, in addition to temporal proximity, the temporal *order* of pre- and postsynaptic spikes is crucial for the sign and magnitude of synaptic modifications (Levy and Steward 1983; Markram et al. 1997; Bi and Poo 1998; Zhang et al. 1998). If the presynaptic neuron A fires an action potential shortly before the postsynaptic neuron B, the synapse from A to B is strengthened. In contrast, if the temporal order of pre- and postsynaptic spike is reversed, the synapse is weakened. The relationship between change in evoked excitatory postsynaptic currents (EPSC) and temporal order of pre- and postsynaptic spikes is shown in Fig. 1.3.

How can this spike timing dependent plasticity rule be used to learn statistical regularities or generate predictions? Abbott and Blum (1996) show in a theoretical study that the temporally asymmetric nature of synaptic modifications predicts interesting modifications in the firing characteristics of populations of neurons encoding information about a stimulus. Specifically, a population of neurons initially representing the current value of a stimulus will, start predicting future values of the stimulus after repeated presentation of the same stimulus sequence. Neurons responding to the stimulus will get activated repeatedly in the same order. This will strengthen synapses from neurons getting activated earlier to those getting activated slightly after. The reverse is true for synapses from later to earlier activated neurons. Thus, after some time neurons will have strong synapses to other neurons which are activated slightly later in the sequence.

Through these strong synapses neurons are getting excited earlier in the sequence, resulting in an overall predictive shift of the population activity.

There are two independent lines of evidence that such predictive population activity does indeed develop with experience. First, Mehta et al. (1997) observed that hippocampal place fields of rats repeatedly traversing the same linear maze shift in the direction opposite to the rat's direction of movement. This effect is observable with as few as 17 trials. Additionally, place fields become negatively skewed, meaning that early spikes undergo larger predictive shifts than spikes fired later during place field traversal (Mehta et al. 2000).

A second line of evidence comes from a series of studies in anesthetized cats' area 17 (V1) carried out in Yang Dan's laboratory (Yao and Dan 2001; Yao et al. 2004; Fu et al. 2002, 2004). Yao and Dan (2001) repeatedly presented neurons with a sequence of two oriented gratings of slightly different orientation (15°). The time between the two gratings was varied from $\pm$8.3 ms to $\pm$41.7 ms. After 1600 such pairings, shifts in preferred orientation towards the orientation presented first were found. Additionally, the magnitude of the shift depended on the temporal separation of the two gratings, in a way resembling the STDP curve shown in Fig. 1.3. The authors interpret this as evidence for spike timing dependent synaptic modifications: Since synapses from neurons preferring the first orientation to neurons preferring the second are selectively potentiated, thus shifting orientation tuning curves towards the orientation presented first. The striking similarity of the time dependence of this effect further supports this claim.

A second study presenting pairs of stimuli in rapid succession at adjacent retinal locations (Fu et al. 2002) shows receptive field shifts which depend in the same way on order and interval between the two stimuli as orientation tuning changes in the previous study. Furthermore, because of complete interocular transfer after monocular conditioning, the underlying changes are very likely caused by cortical mechanisms (Yao et al. 2004).

# 2 Materials and Methods

## 2.1 Electrophysiology

### 2.1.1 Subjects

The experiments were performed on one healthy male adult monkey (macaca mulatta) weighing 11 kg. All experiments were performed at Baylor College of Medicine in Houston, TX (United States of America). The procedures conformed to National Institute of Health (NIH) guidelines and were approved by the Baylor College of Medicine Animal Care and Use Comittee.

### 2.1.2 Surgical methods

Details about the construction of form-specific implants and surgical methods have been published previously (Tolias et al. 2007). Briefly, a cranial head post, a scleral eye coil, a recording chamber, and a chronic tetrode array were implanted under general anesthesia and aseptic conditions. The recording chamber was implanted in the left hemisphere over the operculum in V1 as determined stereotactically based on anatomical high-resolution magnetic resonance images. Fig. 2.1 shows the MRI scan with the targeted recording location highlighted by a white cross.

Due to the form-specific design and high mechanical strength, head post and recording chamber were only secured with screws without the use of dental acrylic or bone cement. After craniotomy and duratomy, the tetrode array was inserted into the record-


Figure 2.1: Anatomical high-resolution magnetic resonance image. Targeted recording location is depicted by the white cross.

ing chamber and it was sealed. Other than the tetrodes themselves nothing else (such as metallic guide tubes) was inserted into the brain.

### 2.1.3 Electrophysiological and eye movement recordings

Neural activity was recorded using custom-built tetrodes whose tips were electroplated with gold. The tetrodes are made of four twisted microwires (Rediohm 800; HP Reid Inc., Palm Coast, FL, USA) with a diameter of 17 $\mu$m each and are independently adjustable. Tetrodes were advanced very slowly (typically <100 $\mu$m/day) on a daily basis for about 2–3 weeks before starting to run experiments. At the time of the recordings, the tetrodes were most likely in supragranular layers of V1. No pre-selection criteria were applied for the neurons recorded. Neural activity was continuously sampled at 32 kHz, digitized and stored to disk using a custom-developed recording system. All processing of neural signals except for a hardware bandpass filter (0.5 Hz to 16 kHz) was done offline as described in section 2.4.1.

During visual stimulation the monkey was required to maintain fixation in a 0.5° radius fixation window. Eye movements were monitored online with the help of an implanted scleral search coil (Judge et al. 1980). These data were collected at 2 kHz and were also stored for offline analysis.

### 2.1.4 Stimulus

Visual stimuli were displayed in a completely dark room on a 22" CRT monitor (Mitsubishi Diamond Pro 2070SB) with a resolution of 1600×1200 pixels (24°×18°) and a refresh rate of 100 Hz. A custom-developed OpenGL/Matlab-based stimulation library which is described in detail in section 2.3 was used to render and display the stimuli.

The stimulus itself was a bright vertically oriented bar (width: 0.15°; height: 5.2°) on gray background. It was either moving continuously in horizontal direction at a speed of 15°/s ("moving bar" condition) or flashed at random locations changing every two frames or 20 ms ("flashed bar" condition).

### 2.1.5 Procedure

During the experiment, the monkey was sitting comfortably in a custom-built primate chair equipped with a head holder to immobilize the head and a juice pump for reward delivery.

Each trial consisted of the following: First, a red fixation spot (0.15° diameter) appeared in the center of the screen on a gray background. Once the monkey had acquired and maintained fixation for 300 ms, the stimulus appeared. While the stimulus was being displayed, he had to maintain fixation for another 500 to 800 ms. Upon successful completion of a trial, the monkey was rewarded with juice which was followed by a short break. If he did not acquire fixation within 3.5 s after the fixation spot appeared or he broke fixation before the stimulus disappeared, the trial was aborted and he did not receive reward.

A complete experiment consisted of an initial receptive field mapping phase (typically 50–150 trials) followed by nine blocks of 100–135 trials each. Such a block contained conditioning phase (85 trials) and a receptive field mapping phase (15–50 trials). In the conditioning phase the moving bar stimulus was played. Direction of motion was

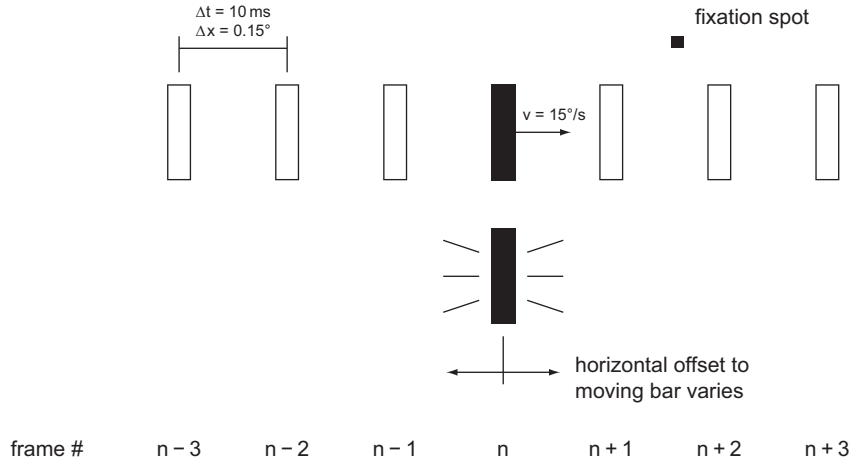


Figure 2.2: Illustration of flash lag effect (FLE) stimulus. The top bar is moving continuously across the visual field while subjects maintain fixation on a red fixation spot. At frame $n$, a second bar is flashed below the moving one for exactly one frame (10 ms). The horizontal offset between the two bars is varied randomly from trial to trial and subjects are asked to report whether the flashed bar was left or right of the moving bar. Flashes were at eccentricities 3.7° and 7.0°. Direction of motion and flash location (left or right of fixation spot) was randomly varied from trial to trial.

selected randomly in each trial according to a probability distribution which was varied during the course of the experiment. In the first three blocks, the bar was moving rightwards and leftwards with equal probability $p = 0.5$; during the next three blocks it was moving rightwards with probability $p = 0.9$ and during the last three blocks with $p = 0.1$. In the receptive field mapping phases the flashed bar stimulus was played for 500–800 ms in order to map one-dimensional receptive fields.

## 2.2 Psychophysics

### 2.2.1 Subjects

Subjects for the human psychophysical study were three individuals (the author and two graduate students recruited from our laboratory). In the monkey psychophysics study one healthy male adult monkey (macaca mulatta) participated.

### 2.2.2 Stimulus

The stimulus is illustrated in Fig. 2.2. A vertically oriented bar is moving horizontally across the screen. At some point a second bar is flashed below it for one frame (10 ms). Location of the flashed bar and horizontal offset between the two bars were varied randomly from trial to trial. The subjects' task was to indicate whether they perceived the flashed bar left or right of the moving bar.

Moving and flashed bars had the same size (width: 0.27°; height: 4.0°). The moving bar was moving on a horizontal trajectory of 21.4° centered around the fixation spot at a speed of 13.4°/s. The vertical distance (edge to edge) between moving and flashed bar was 0.80°. The flashed bar was flashed at two different locations (eccentricity of flashed

bar center: 3.7° on either side of the fixation spot; moving bar: 1.9°) and with offsets of 0°, ±0.27°, ±0.80°, ±1.61°, and ±2.41° relative to the moving bar.

### 2.2.3 Procedure

The stimulus was presented on a 22″ CRT monitor (Mitsubishi Diamond Pro 2070SB) at a resolution of 1600×1200 and a refresh rate of 100 Hz using the visual stimulation library (section 2.3) for rendering and display.

In the human study subjects were asked to fixate on a red fixation spot (0.13° diameter) in the center of the screen and move their head as little as possible. In a two-alternative forced choice task, they responded by moving a joystick in the direction in which they perceived the flashed bar relative to the moving bar.

In the monkey study the same general procedure as in the electrophysiology experiments was used: the trial was started by a fixation spot appearing and only after the monkey had acquired fixation, the stimulus was played. After the stimulus had completely disappeared, the monkey had to make a response by moving a joystick in one of two directions, just as in the human study. Rewarding, however, is non-trivial in this case; we will outline our approach here.

The basic problem in rewarding the monkey in this study is the dissociation between physical reality and perception. On the one hand, the monkey has to be rewarded for fixating and making correct joystick movements in order to train him the task. On the other hand, he is supposed to indicate whether he *perceived* the flashed bar left of right of the moving bar. It is the monkey's percept which is being studied and—in contrast to the physical reality—is unknown to the experimenter. Thus, when the offset between moving and flashed bar is sufficiently small, it is impossible to know if the monkey actually reported his percept. To solve this problem we split the trials in *supra-threshold* and *probe* trials. In supra-threshold trials, the horizontal offset between flashed and moving bar was well above the magnitude of the flash lag effect as estimated in the human study (see section 3.1.1). During training, we only used supra-threshold trials with offsets between 2° and 4° so that there was always a well-defined notion of a correct or incorrect response, independent of whether monkeys perceive the illusion or not. Until he had learned to perform the task correctly (> 90% correct responses), the monkey was only shown supra-threshold trials and rewarded for correct responses.

In probe trials, reward cannot carry any information about the true offset, for otherwise the monkey would be trained the response one is later misinterpreting as his perception. However, simply rewarding a fixed percentage (such as none, all, or 50%) of probe trials irrespective of the response is dangerous. If the monkey learns this contingency, he is very likely to start responding in an unwanted way such as always giving the same answer. Therefore, we chose a slightly more complicated approach concealing this contingency. During training, we habituated him on only receiving reward after having correctly completed a randomly chosen (between one and three) number of trials. During the experiment we then inserted probe trials to test the monkey's percept. Probe trials were never counted towards reward and, consequently, could not re-enforce any particular choice of response. As the monkey was accustomed to having to complete multiple trials before getting reward, there was strong motivation for him also to report his percept in probe trials. Furthermore, he could not infer a potentially correct response to a probe trial since there are two different possible explanations for not getting reward

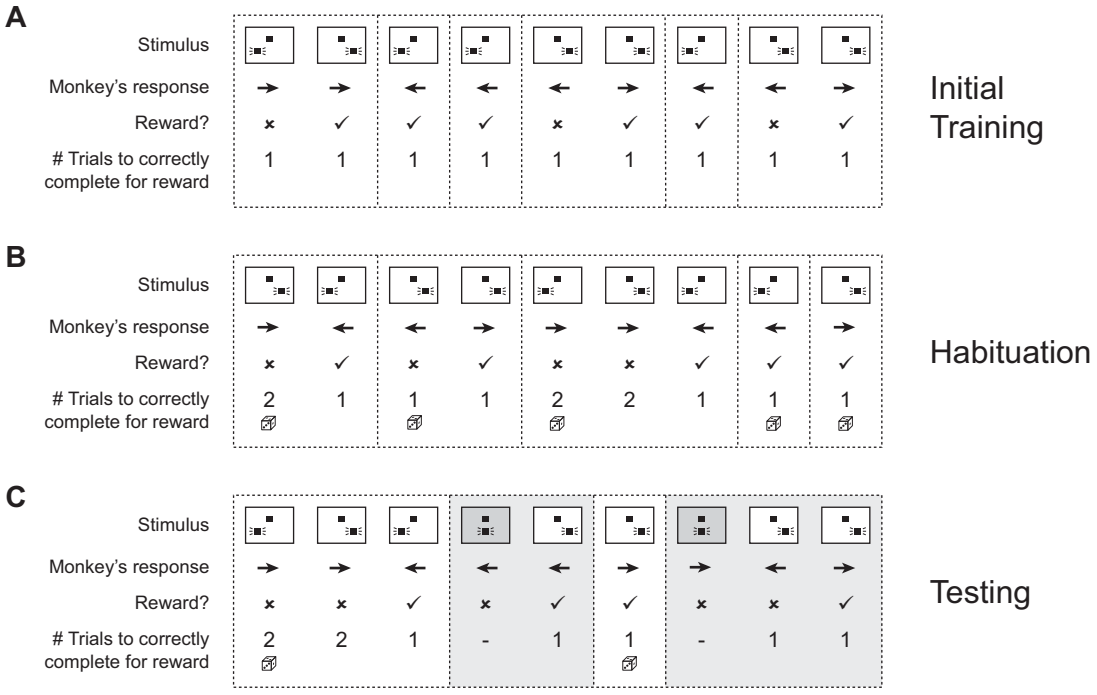


Figure 2.3: Illustration of rewarding scheme. Three phases: **A:** Initial training. Each column represents one trial. Top row depicts the stimulus. The top square is the moving target, the bottom square the flashed one. Second row indicates the monkey's response. The monkey has to report the location of the flashed relative to the moving. Third row indicates whether the monkey was rewarded for this trial. Fourth row indicates the number of trials to be completed correctly until reward is given. During the initial training, every correct trial was rewarded. **B:** Habituation phase. In this phase we started habituating the monkey to only being rewarded after completing $n$ trials correctly, where $n$ was chosen randomly between 1 and 2 at the first trial following reward delivery (depicted by the die icon below the number). **C:** Test phase. Here we insert probe blocks. Always the first trial in a probe block was the probe trial and the monkey had to complete another trial after the probe trial with the correct response in order to get rewarded, irrespective of his response in the probe trial. Probe blocks are highlighted by gray background; probe trials are depicted by horizontally aligned squares, although there might have been a small offset below threshold.

in a probe trial and then getting reward after the next: In the first scenario, the number of trials to be correctly completed for reward was two and the monkey responded correctly in both trials (Fig. 2.3 B, first block). In the second scenario, the number of trials to be correctly completed was one and the monkey responded wrong in the first trial (Fig. 2.3 B, second block). Thus, the monkey cannot draw any conclusions about whether he responded correctly or not in the probe trial. The rewarding scheme and its three phases are illustrated in Fig. 2.3.

### 2.2.4 Data analysis

For fitting of psychometric functions and obtaining confidence intervals the free, open-source Matlab-based psignifit toolbox (Wichmann and Hill 2001a, b) was used. First,

observers' responses were converted from left/right to leading/lagging (where leading means the response was right in a rightwards motion trial or left in a leftwards motion trial). Then, a logistic function

$$\psi(x) = \gamma + (1 - \gamma - \lambda)F(x) \tag{2.1}$$

where

$$F(x) = \left(1 + \exp\left(\frac{a - x}{b}\right)\right)^{-1} \tag{2.2}$$

was fitted to the fraction of responses indicating the flash leading the moving bar at different offsets $x$. Here, $\gamma$ and $\lambda$ account for stimulus-independent lapses of observers while $a$ and $b$ determine slope and position of the psychometric function.

The quantity of interest in this study is the point of subjective equality (PSE), which is the offset where flashed and moving bar are perceived to be aligned. This point is taken to be the offset $x$ where $F^{-1}(x) = 0.5$, i. e. the value where observers do not respond in favor of any of the two directions. The maximum-likelihood estimate of the PSE is given by $a$ and confidence intervals are obtained by parametric bootstrap (Wichmann and Hill 2001b).

## 2.3 Visual stimulation library

### 2.3.1 Goals

As a part of this project a library of classes to facilitate development of visual stimuli for psychophysical or electrophysiological studies was developed. We identified two main requirements for such a library:

1. Accurate timing. Whenever responses (by neurons or subjects) to a visual stimulus are studied, it is of particular importance to precisely control onset and offset of the stimulus on a millisecond precision.

2. Data storage. Usually, visual stimuli are parameterized in many different ways, stimulus conditions are randomized across trials, and various sorts of events might be triggered and saved during each trial. All these data have to be stored in a consistent way so the experimenter can analyze the data efficiently and, if necessary, reconstruct the course of the entire experiment.

Further, the system is intended to only act as a client processing requests (such as show stimulus, save data, etc.). Control of the experiment, which involves processing of analog data to react to eye and joystick movements or controlling hardware for reward delivery, is handled by a custom-developed LabView-based state system (R. James Cotton; Tolias Lab).

### 2.3.2 Implementation

The library is implemented in Matlab using the free open-source Psychophysics Toolbox (Brainard 1997). The Psychophysics Toolbox was chosen as the basis since it provides a convenient OpenGL interface to Matlab and, most importantly, very accurate timing of graphics output. The developed library contains five basic classes which we will briefly describe here.

`BasicExperiment`. This is the base class for every stimulus irrespective of what type (trial-based, continuous). It processes function calls received via network communication from the state system. Furthermore, it provides functionality to synchronize timestamps of the stimulation system to the state system and provides the basic drawing window for visual stimuli.

`TcpConnection`. This class is an abstraction of the network connection. It parses incoming TCP packets and writes data to the network according to a defined network API shared by the state system.

`PhotodiodeTimer`. This class is used to maintain highly accurate timestamps about the visual display. It keeps track of every single graphics card buffer swap by storing a high-precision timestamp obtained from the Psychophysics Toolbox in an internal buffer. Additionally, the color of a small square spot in the top-left corner of the monitor is alternated between black and white at every buffer swap, making it possible to also record buffer swaps externally via a photodiode attached to the monitor. Although there is special purpose graphics hardware available which sends a TTL pulse on every buffer swap, these cards are typically both very expensive and not state of the art in terms of rendering performance. We therefore favored this software/photodiode based approach which was found to create no programming overhead when developing stimuli and work reliably. Furthermore, comparison of timestamps obtained from the Psychophysics Toolbox to the photodiode signal confirmed the PTB timing accuracy up to a few tens of $\mu$s.

`TrialBasedExperiment`. This class inherits `BasicExperiment` and implements the main tasks necessary to run a trial-based experiment. This includes start and end of trial/session, drawing fixation spot, playing sounds, logging events, and storing/retrieving parameters. For a simple stimulus the only necessary additional code to implement is the function that draws the stimulus. By inheriting from this class and overwriting functions, stimuli can be highly customized.

`StimulationData`. A class that holds all parameter values and events for every trial of a trial-based experiment. It provides a simple interface to write or retrieve parameter values and log events. It is saved as a Matlab data structure at the end of the experiment making it very easy to use the stored data for analysis without the need of additional file parsing. Additionally, after every trial, data is saved incrementally to avoid losing data due to a system crash during the experiment.

## 2.4 Data analysis

All analysis was done in Matlab (The Mathworks Inc., Natick, MA) using custom-written software. Analysis programs are based on a previously developed data management system (Ecker et al. 2007a, b), which can be used to organize, store, and retrieve both processed and raw neural data.

### 2.4.1 Single unit extraction

#### 2.4.1.1 Spike detection

In this study, we are interested in characterizing how neurons fire action potentials (or spikes) in response to a certain stimulus and how these firing characteristics change. Since the neural signal is recorded as a continuous broadband signal, the first step is to detect spikes in the recorded signal. For this purpose, the signal is first bandpass filtered to 600–6000 Hz. After filtering, the noise level is estimated and all events exceeding a certain threshold are considered action potentials.

For filtering we use an equiripple linear phase FIR filter with a passband of 600–6000 Hz. The filter is designed using the Parks-McClellan algorithm (Digital Signal Processing Committe 1979) provided by Matlab's Signal Processing Toolbox. The don't care bandwidth is 200 Hz, attenuation outside the passband 65 dB, and maximum passband ripple 0.002 dB.

For robust and consistent spike detection it is useful to automatically determine the detection threshold from the noise distribution of the data. We observed that the bandpass filtered signal was approximately normally distributed. However, especially if a lot of action potentials were recorded by a given tetrode, the distribution tends to have much heavier tails than predicted by a normal distribution. In order to avoid a bias introduced by the number of spikes, the noise standard deviation $\sigma_n$ is estimated by cap fitting (Thakur et al. 2007). In cap fitting, only a limited region around the origin is fitted by a Normal distribution ignoring the tails of the distribution. Given this robust estimate of the noise level, any event exceeding $5\sigma_n$ on any of the four tetrode channels is considered an action potential.

After an action potential has been detected, it is re-aligned to its peak in order to correct for jitter induced by noise and the discrete triggering mechanism. To do so, a segment of 1 ms (32 samples) is extracted and up-sampled 10-fold by cubic spline interpolation. This up-sampled segment is then aligned to its center of mass $c = \frac{1}{n}\sum_{i=1}^{n} ix_i$, down-sampled to the original sampling rate and stored to disk.

#### 2.4.1.2 Spike sorting

Extracellular electrodes such as the tetrodes used in this study typically record many different neurons at the same time. The goal of spike sorting is to identify clusters of action potential that were emitted by single neurons and separate those from background activity which cannot be assigned to any specific neuron due to its low amplitude. This background activity is typically called *multi unit activity* and will also be considered in this study.

While spike sorting is a rather difficult problem with single electrodes and they have to be positioned closely to the neuron(s) being recorded, tetrodes usually do not need to be positioned as carefully and can potentially separate more neurons. The reason for this is that they consist of four wires (channels) whose tips are only spaced a few $\mu$m. Thus, most neurons are recorded on all four channels. Depending on where each neuron is located relative to the tetrode, its recorded signal will be stronger on the channels closer to the neuron than on the channels further away. This allows separation of different neurons even if their signal on one specific channel is almost identical, a fact that is also illustrated in Fig. 2.4.

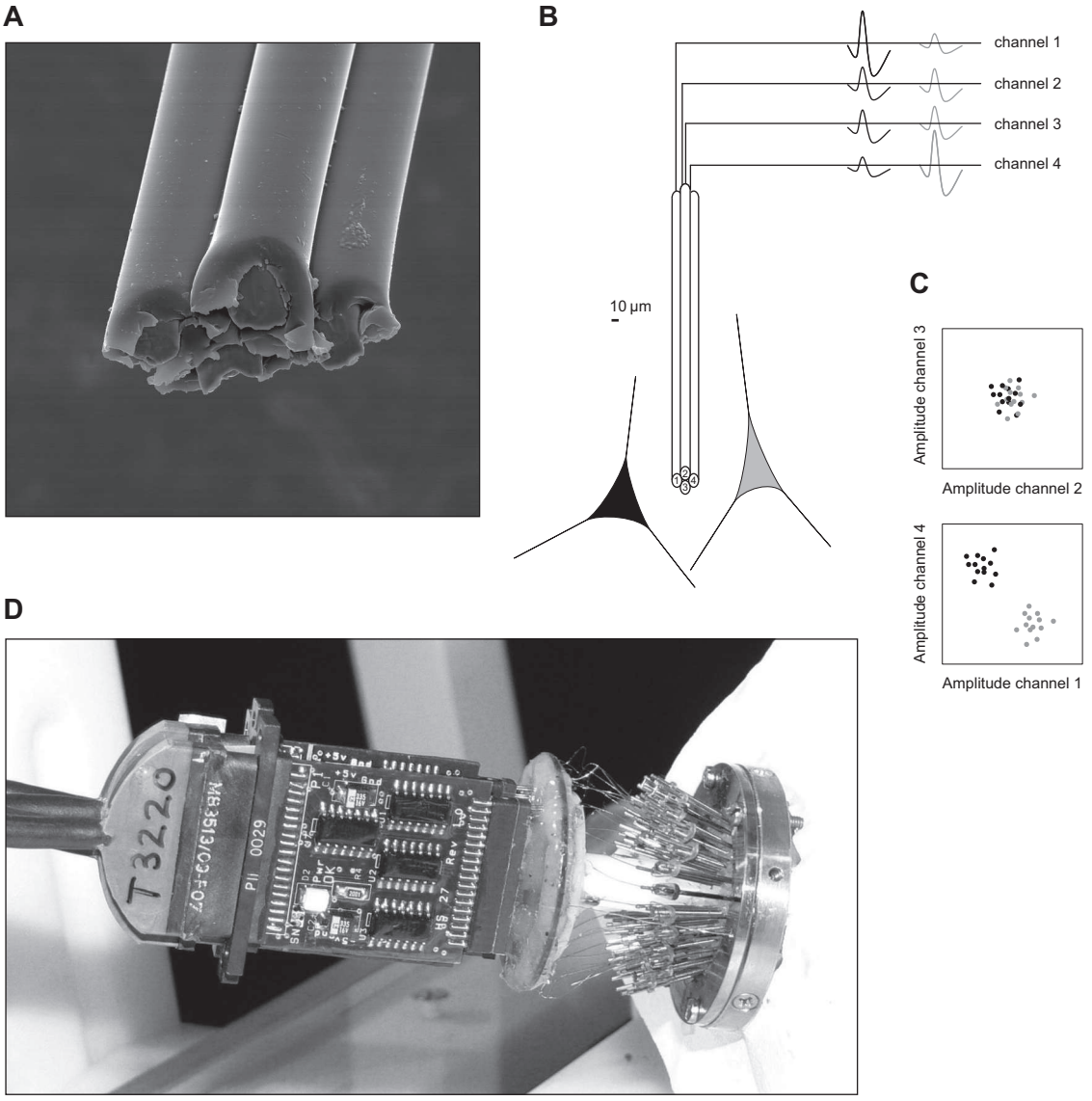


Figure 2.4: **A:** Electromicroscopy photo of a tetrode. **B:** Schematic illustration of a tetrode with two neurons and the recorded waveforms on each of the four tetrode channels. Because the black neuron is closest to channel 1, its amplitude on this channel is larger than on the others. The gray neuron, however, is closer to channel 4 and therefore has its largest amplitude on channel 4 and smallest on channel 1. **C:** Scatter plot of waveform amplitudes of multiple action potentials of the two neurons recorded on different tetrode channels plotted against each other. The advantage of tetrodes over single electrodes can be seen when considering only channels 2 and 3: Since both neurons have approximately equal distance to those channels, they produce waveforms with very similar amplitudes, making single unit isolation particularly hard. In contrast, because the two neurons have different distances to channels 1 and 4, on those channels they can clearly be distinguished. Because of their four channels tetrodes always triangulate each neuron's position and yield better single unit isolation. **D:** Recording drive holding 24 tetrodes. On the right side the titanium chamber implanted on the monkey's head. Attached to it the guide screws for the independently adjustable tetrodes. Tetrodes are the thin wires connected to the round plate which the pre amplifiers (chips in the left half) are connected to. Scale approximately 1:1.

The methods used for spike sorting have been published previously (Tolias et al. 2007), so we only briefly discuss them here. First, for dimensionality reduction, principal component analysis (PCA) is performed on the extracted spike waveforms on a per-channel basis. The first three principal components on each channel are used, resulting in a 12-dimensional feature vector for each spike. Then, a mixture of Gaussians model given by

$$P(\boldsymbol{x}) = \sum_{i=1}^{n} \pi_i \frac{1}{K_i} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_i)^T \boldsymbol{Q}_i^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_i)\right) \tag{2.3}$$

is fitted to the data using the split & merge expectation maximization algorithm (Ueda et al. 2000). The mixture model consists of $n$ clusters with means $\boldsymbol{\mu}_i$ and covariance matrices $\boldsymbol{Q}_i$; $\pi_i$ is the prior probability of class membership. In other words, $\pi_i$ is the fraction of the total number of spikes that was generated by cluster $i$. Each spike is then assigned to the cluster that maximizes the posterior probability

$$\begin{aligned} P(i|\boldsymbol{x}) &= \frac{P(\boldsymbol{x}|i)P(i)}{P(\boldsymbol{x})} && (2.4)\\ &\sim \pi_i \frac{1}{K_i} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_i)^T \boldsymbol{Q}_i^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_i)\right). && (2.5) \end{aligned}$$

#### 2.4.1.3 Single unit isolation quality

When analyzing the behavior of extracellularly recorded single neurons it is imperative to quantify how well the spike sorting procedure isolated each neuron and what the probability of falsely assigned spikes is. The fact that our spike sorting procedure returns a parametric model of the data can be exploited to estimate the quality of single unit isolation in a rather straightforward and well-defined way.

We sample from the probability distribution defined by the mixture model to generate a synthetic data set. Using the same classification function as the clustering algorithm (see section 2.4.1.2 above), every data point is assigned to one the $n$ clusters. Since in the synthetic data set the generating components are known for all data points, false positive and miss rates can be estimated. For further details on this procedure we refer to Tolias et al. (2007).

### 2.4.2 Receptive field mapping

Linear receptive fields were mapped by computing the spike-triggered average response of each neuron to a vertically oriented bar (width: 0.15°; height: 5.2°) at locations randomly chosen along the $x$-axis and changing every second monitor frame (or 20 ms). No spatial and temporal smoothing was applied. The receptive field was defined as the first Wiener kernel, which is given by

$$D(x,\tau) = \frac{\langle r \rangle C(x,\tau)}{\sigma_s^2} \tag{2.6}$$

and where $C(x,\tau)$ is the spike-triggered average stimulus, $\langle r \rangle$ the neuron's mean firing rate, and $\sigma_s^2$ the variance of the stimulus (Dayan and Abbott 2001). To compute the

spike-triggered average, the stimulus $s(x,t)$ in a window $[t_i - \tau_{max}, t_i]$ preceding each spike at time $t_i$ is extracted and averaged across all spikes:

$$C(x,\tau) = \frac{1}{n}\sum_{i=1}^{n} s(x, t_i - \tau), \tag{2.7}$$

where $n$ is the total number of spikes. Typically, $D$ was mapped for $x$ in the range of $\pm 1^\circ$ around the location of the receptive field and $\tau_{max} = 150$ ms. In this case, a positive $\tau$ corresponds to the influence of the stimulus at time $t - \tau$ on the firing rate at time $t$.

To obtain an estimate of the spatial and temporal size and location of each receptive field, we make the assumption that the receptive fields are separable (i. e. can be written as $D(x,\tau) = D_s(x)D_t(\tau)$) and fit the spatial and temporal components $D_s(x)$ and $D_t(\tau)$ independently by a parametric function. The best separable approximation $\hat{D}$ to $D$ (in a least squares sense) can be obtained by computing the singular value decomposition (SVD)

$$D = USV^T \tag{2.8}$$

and setting $\hat{D}_s = \sqrt{s_1}\boldsymbol{u}_1$ and $\hat{D}_t = \sqrt{s_1}\boldsymbol{v}_1$, resulting in $\hat{D} = \hat{D}_s\hat{D}_t = \boldsymbol{u}_1 s_1 \boldsymbol{v}_1$. Here, $\boldsymbol{u}_1$ and $\boldsymbol{v}_1$ are the first columns of $U$ and $V$, respectively, and $s_1$ is the largest singular value of $D$. We then fit a scaled normal distribution with an offset added given by

$$f(x) = f_0 + a\exp\left(-(x-\mu)^2/\sigma^2\right) \tag{2.9}$$

to each component using least squares. $\mu$ is used as an estimate for the center of the receptive field and the size is defined to be $2\sigma$.

Since, due to orientation tuning, some of the receptive fields are inhibitory, we define the temporal component $D_t$ to have a positive peak. For the spatial component $D_s$ we fit both $f(x)$ and $-f(x)$ and use the fit that has the smaller residual error to decide whether it is inhibitory or excitatory.

For the analysis in section 3.3 only a very limited number of trials was available to estimate the receptive fields. Since the SVD-based method was found to only work reliably with a sufficiently high number of spikes, we chose a slightly different approach to estimate receptive field centers for subblocks of the experiment. We first collapsed all trials and extracted the parameters as described above. Then, for each individual subblock $i$, we extracted the central portion of the (noisy) receptive field for this block by taking a region of $\pm 3\sigma_x$ and $\pm\sigma_t$ around the center. To obtain an estimate of the spatial RF component $D_s^{(i)}$ of subblock $i$, we averaged this region over time:

$$D_s^{(i)}(x) = \frac{1}{2\sigma_t}\int_{\mu_t-\sigma_t}^{\mu_t+\sigma_t} D(x,\tau)d\tau \tag{2.10}$$

The result was further smoothed by a Gaussian window of $0.05^\circ$ s.d., which typically resulted in relatively smooth and reliable spatial receptive field profiles. These were then fitted by the scaled normal distribution function defined above (Eq. 2.9) and $\mu$ was used as an estimate of the center.

Skewness was computed as the third standardized moment according to

$$\gamma = \frac{1}{\sigma_x^3 K}\int_x D_s^{(i)}(x)(x-\mu)^3 dx \tag{2.11}$$

where

$$K = \int_x D_s^{(i)}(x)dx \tag{2.12}$$

is the normalization constant to make $D_s^{(i)}$ a properly normalized distribution and integrals were in practise truncated at $\pm 3\sigma_x$ around the center.

#### 2.4.2.1 Constructing linear prediction of firing rates

Given the receptive field $D(x, \tau)$ the optimal linear prediction $r_{\text{pred}}$ of the firing rate $r$ in response to stimulus $s(x, t)$ can computed by (Dayan and Abbott 2001)

$$r_{\text{pred}} = \int_x \int_\tau D(x, \tau) s(x, t - \tau) d\tau dx \tag{2.13}$$

### 2.4.3 Response latency estimation

Response latency was defined as the time between the stimulus appearing in the center of a neurons receptive field and the neural response reaching 50% of its peak firing rate. Taking the spike density functions for both directions of motion, the response latency to the moving bar can be computed without knowledge of shape or position of the neuron's receptive field. The reason for this is simple: The bar is moving along the exact same trajectory of length $x_t$ in the two conditions, but in opposite directions. It is further moving at constant speed $v$. Thus, if the receptive field center is at $x_{\text{rf}}$, in one condition it takes the bar $t_{\text{right}} = x_{\text{rf}}/v$ to reach the RF center; in the other condition it takes the bar $t_{\text{left}} = (x_t - x_{\text{rf}})/v$. Taking the two directions together, it takes the bar a total of $t_{\text{total}} = x_t/v$ to reach the RF center. Therefore, the latency can be computed as

$$\tau_{\text{moving}} = \frac{t_{\text{onset,right}} + t_{\text{onset,left}}}{2} \tag{2.14}$$

where $t_{\text{onset}}$ is the time from stimulus onset to the spike density function reaching 50% of its peak value for the respective direction of motion.

# 3 Results

## 3.1 Psychophysics: flash lag effect

### 3.1.1 Human study

The human study served as a pilot study to demonstrate that human subjects perceive the flash-lag illusion under stimulus conditions similar to those later used in the monkey psychophysics and electrophysiology experiment. Due to the plethora of different parameters influencing the magnitude (and some even the sign) of the illusion, it also helped to obtain an estimate of what effect magnitude will be expected in the monkey study. Given that visual processing is generally somewhat faster in monkeys than in humans (Boch and Fischer 1986), a second goal of this study was to make sure the chosen parameters result in an effect large enough to be detected in potentially reduced form in the monkey.

A standard flash-lag setup in the continuous motion condition was used. While subjects were fixating on a fixation spot in the center of the screen, a bar was moving horizontally across the visual field. At some point during the course of the movement (at ~1.5° eccentricity from the fovea), a second bar was flashed below the moving one for exactly one frame. The offset between flashed and moving bar was varied randomly from trial to trial and subjects were asked to report by means of moving a joystick whether the flashed bar was left or right of the moving bar. For further details see methods, section 2.2.

Fig. 3.1 A shows the psychometric function of one observer (P. B.). The fraction of trials in which this observer reported the flashed bar to lead the moving bar are plotted as a function of the horizontal offset between the two bars at the time of the flash. Clearly, the curve is shifted to the right, showing the presence of a flash-lag effect. In fact, in ~90% of the trials in which the flashed bar was leading by 0.27° ahead of the moving bar at the time of the flash, this observer reported it to be lagging behind. The point of subjective equality (PSE; i. e. where the to bars are perceived to be aligned) was found to be at 0.82° (95% c. i.: 0.49° < PSE < 1.03°) for this observer. The PSE for all three subjects that participated in this study are shown in Fig. 3.1 B. Error bars indicate 95% confidence intervals, derived by bootstrap methods (Wichmann and Hill 2001b). While for all three subjects there is a highly significant flash-lag effect, the magnitude of the effect is quite different between observes. For observer A. E. (author) the PSE is 0.48° (0.31° < PSE < 0.63°) and for observer A. H. it is 0.88° (0.68° < PSE < 1.02°).

### 3.1.2 Monkey study

If one is interested in the neurophysiological basis of the flash-lag illusion, it is obviously of particular importance to know whether or not certain animal species also perceive the flash-lag illusion. We hypothesized that given the analogy between the monkey and human visual system, monkeys should also perceive the flash-lag illusion.

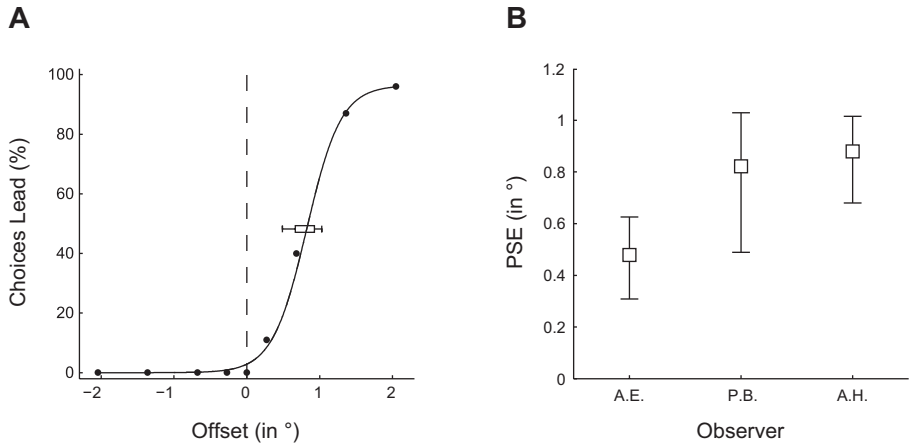


Figure 3.1: Human FLE results. **A:** Psychometric function for observer P. B. The fraction of trials in which the flashed bar was reported to lead the moving bar is plotted as a function of the horizontal offset between the two bars. Point of subjective equality (PSE) is depicted by the horizontal box. Size of box corresponds to 68%, whiskers depict 95% confidence intervals. **B:** Magnitude of mislocalization for all three subjects (including the author). Error bars indicate 95% confidence intervals estimated by parametric bootstrap.

To test this hypothesis, we trained a monkey to report the location of a flashed bar relative to a moving bar—the same way human subjects are instructed to do in a typical flash-lag experiment. For rewarding we used a special rewarding scheme allowing to probe the monkey's percept without re-enforcing any particular choice of answer. This ensured that rewarding did not create a bias in the monkey's responses which could be falsely interpreted as a flash-lag effect. Briefly, the offset between moving and flashed bar was varied randomly between 0° to 2.46°. Trials with offsets of less than 1.5° were considered *probe trials* and did not count towards rewarding. Trials with offsets of at least 1.5° (approximately twice the PSE of the average human subject in the human study) were considered *supra-threshold* and were rewarded if the monkey indicated the right location (see methods for details). Note that this rewarding scheme results in a ratio of approximately 1:4 probe trials to supra-threshold trials, requiring a lot of data to obtain good estimates for the psychophysically interesting offsets.

We ran the experiment at two different contrast levels (Michelson contrast 0.30 and 0.43 (Michelson 1927)) and collected a total of 1287 and 930 trials for low and high contrast, respectively. This resulted in an average of 42 and 33 trials per offset below threshold and 269 and 191 trials per offset above threshold for high and low contrast, respectively.

The monkey's psychometric functions at the two contrast levels are shown in Fig. 3.2 A and B. In both, a clear flash-lag effect can be observed. The monkey's PSE is 0.43° (95% c. i.: 0.26° < PSE < 0.60°) at low contrast and 0.49°(0.34° < PSE < 0.64°) at high contrast. Although this is comparable to the human study, it should be mentioned that in the human study, a contrast of 0.17 was used. This change of parameters was necessary as the monkey's overall performance (correct answers in supra-threshold conditions) was relatively poor at the low contrast.

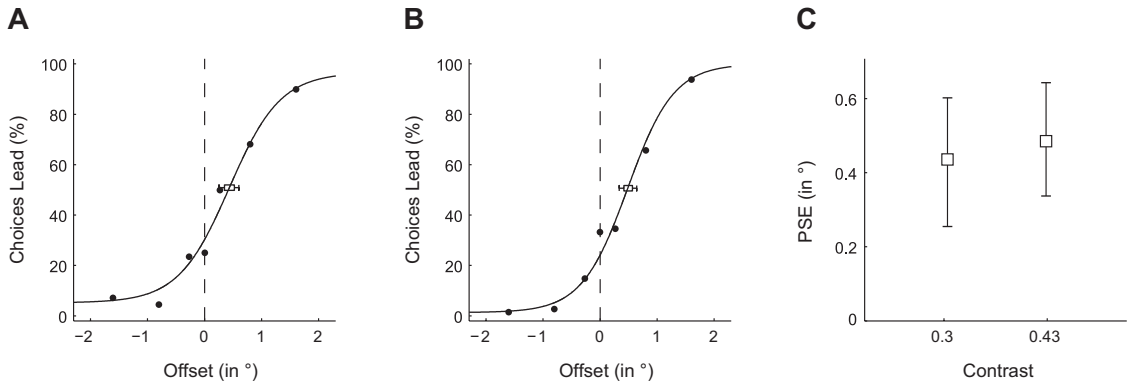


Figure 3.2: Monkey FLE results. **A:** Psychometric curve for low contrast (Michelson contrast 0.30). The fraction of trials in which the flashed bar was reported to lead the moving bar is plotted as a function of the horizontal offset between the two bars. Point of subjective equality (PSE) is indicated by horizontal box. Width of box depicts 68%, whiskers 95% confidence intervals. **B:** As in A, but for high contrast (0.43). **C:** PSE for the two contrasts. Error bars indicate 95% confidence intervals. PSE are 0.43° for low contrast and 0.49° for high contrast (not significantly different, parametric bootstrap)

## 3.2 Differential latencies in V1

The psychophysical study presented in the previous section confirmed our hypothesis that monkeys also perceive the flash lag illusion. We were now interested whether we can find any electrophysiological evidence potentially explaining the perception of this illusion in area V1. Among the various hypotheses introduced earlier, the one that makes a clear prediction about the neural basis of the flash-lag effect is the differential latencies hypothesis. In this section, we seek to address the question whether neurons in primary visual cortex respond faster to moving bars than to flashed ones—as predicted by the differential latencies model.

To this end, we carried out an electrophysiological study using a stimulus that involves both moving and flashed bars. The stimulus and its various conditions are described in detail in sections 2.1.4 and 2.1.5. Here we focus on the distinction between moving and flashed bar conditions and ignore the prior probability of motion direction for the moment. We will, however, come back to this point later in sections 3.3 and 3.4. We first studied the spatial and temporal structure of the receptive fields of single neurons and multi unit activity using the flashed bar stimulus. Then, we compared response latencies between flashed and moving stimulus in order to answer the question whether there are differential latencies in V1.

### 3.2.1 Linear receptive fields

Receptive fields of 55 single neurons and 131 multi units from the superficial layers of V1 were mapped using reverse correlation (see methods section 2.4.2 for details). Vertically oriented bars of light were flashed at random locations along the $x$-axis. From this, a two-dimensional firing profile can be extracted, which is a function of both location of the flash and time relative to the flash. Fig. 3.3 shows examples of such space/time receptive fields for both single unit and multi units.

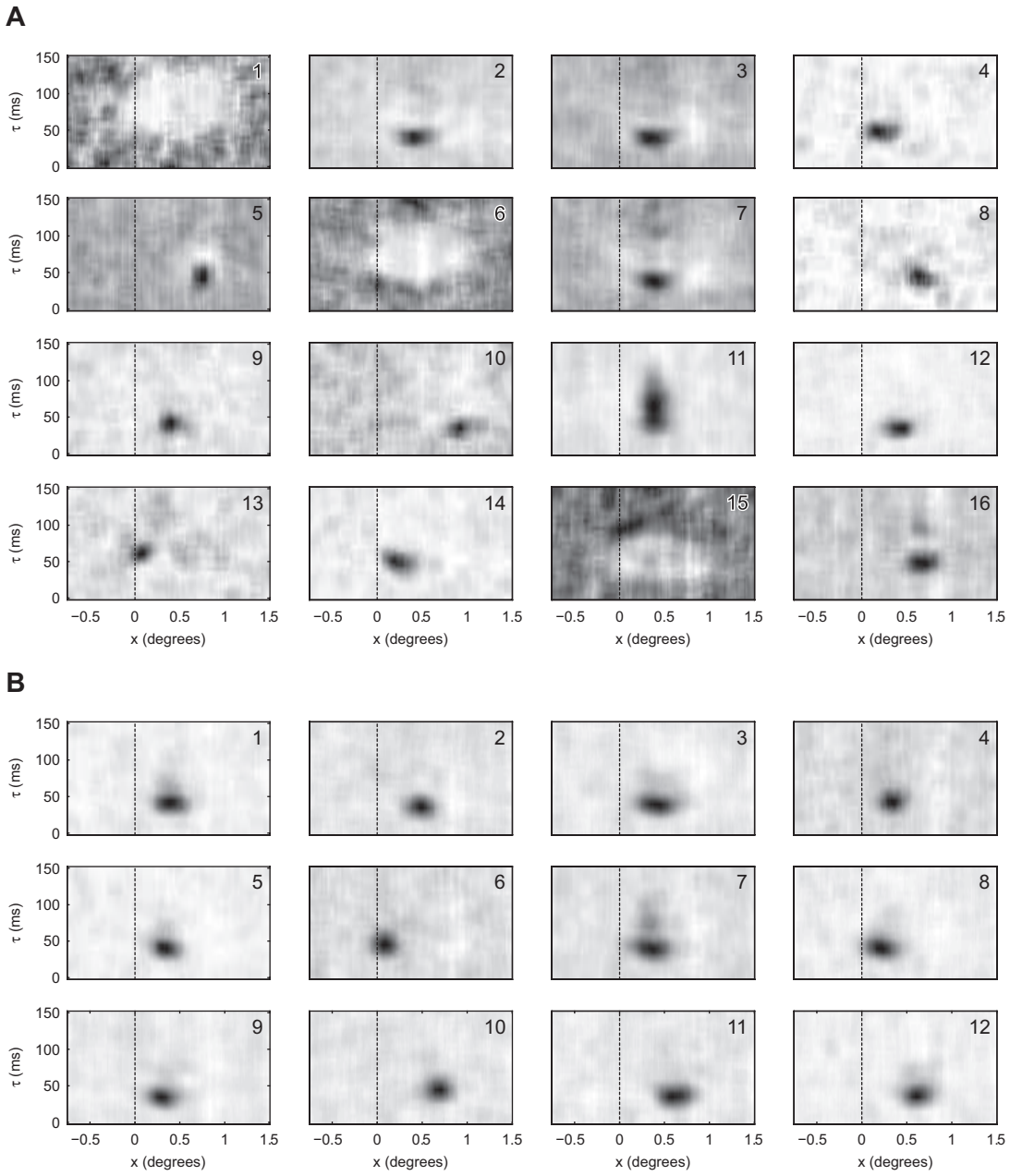


Figure 3.3: Linear receptive fields obtained by reverse correlation. **A:** Single unit receptive fields. Some of the cells (1, 6, and 15) are inhibited, potentially because of orientation tuning or extra-classical receptive field effects such as end stopping (Hubel and Wiesel 1968; Bolz and Gilbert 1986). **B:** Multi unit receptive fields. In contrast to the single units, multi unit receptive fields are much less diverse. Most likely this is because they are averages of many neurons' receptive fields.

To quantify receptive field sizes and locations, spatial and temporal components were obtained via singular value decomposition (SVD). Both spatial and temporal component were parameterized by fitting a normal distribution (see methods for details). Receptive fields were found to be located in the lower right hemifield at horizontal eccentricities ranging from about 0.1° to 0.9° for both single and multi units. This is consistent with

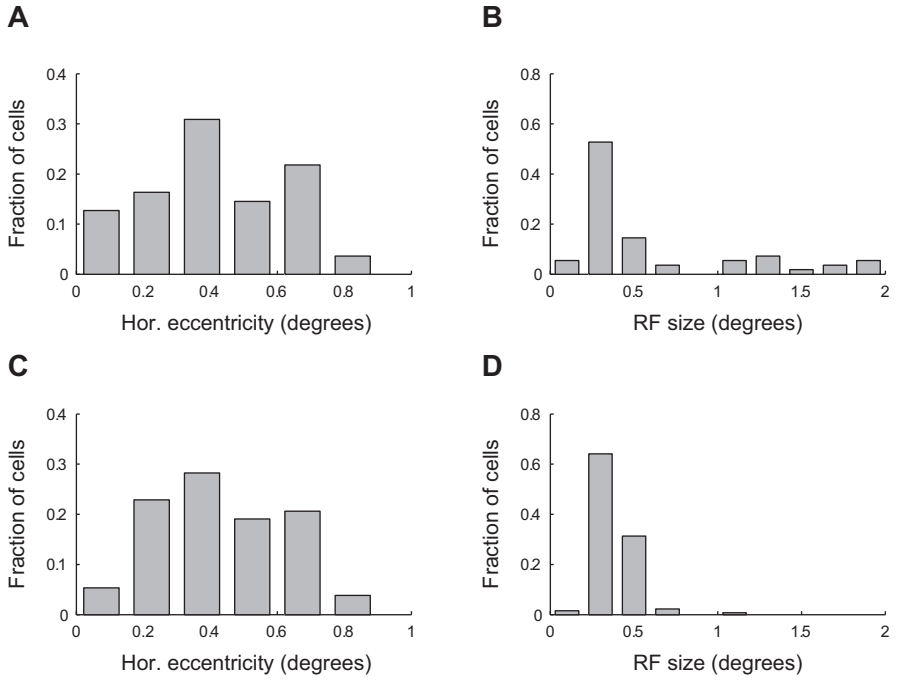


Figure 3.4: Spatial receptive field (RF) statistics. **A:** Distribution of single unit RF centers. Peak of the parametric fit to the spatial component of the RF was used (see methods in section 2.4.2 for details) **B:** Distribution of single unit RF sizes. Sizes were estimated from the parametric fit and defined as $2\sigma$. **C:** Distribution of multi unit RF centers computed as in A. **D:** Distribution multi unit RF sizes computed as in B.

the targeted recording site over the operculum and 2–5 mm posterior to the lunate sulcus (Gattass et al. 1981). Single units were on average located at an eccentricity of $0.41^\circ$ with a standard deviation of $0.21^\circ$. Multi units were found at $0.43 \pm 0.18^\circ$. The full distributions of receptive field centers are shown in Fig. 3.4 A and B. Receptive fields were, on average, $0.36 \pm 0.17^\circ$ wide (where width was defined as two times the standard deviation of the parametric fit) for single units excited by the stimulus and $1.14 \pm 0.70^\circ$ for inhibited ones. Multi unit receptive fields were comparable to those of excited single units and spanned $0.39 \pm 0.10^\circ$.

### 3.2.2 Comparison of response latencies for moving vs. flashed bars

We now come back to the initial question asked at the beginning of this section: how do response latencies to flashed and moving bars compare? It is conceivable that due to the predictive nature of the moving stimulus, neurons can respond faster to moving bars than to flashed bars, giving rise to illusions like the flash lag effect.

For the moving bar stimulus, we defined response latency as the time between the stimulus appearing in the center of a neuron's receptive field and the firing rate reaching 50% of its peak rate (for details, see methods in section 2.4.3). For the flashed bars, latencies were extracted from the temporal component of the receptive field. For the moving bars, the spike density function was used and latencies for the two directions of motion were averaged. Motivated by the observation of the previous section that inhibitory receptive fields are significantly larger (both in space ant time) than excitatory

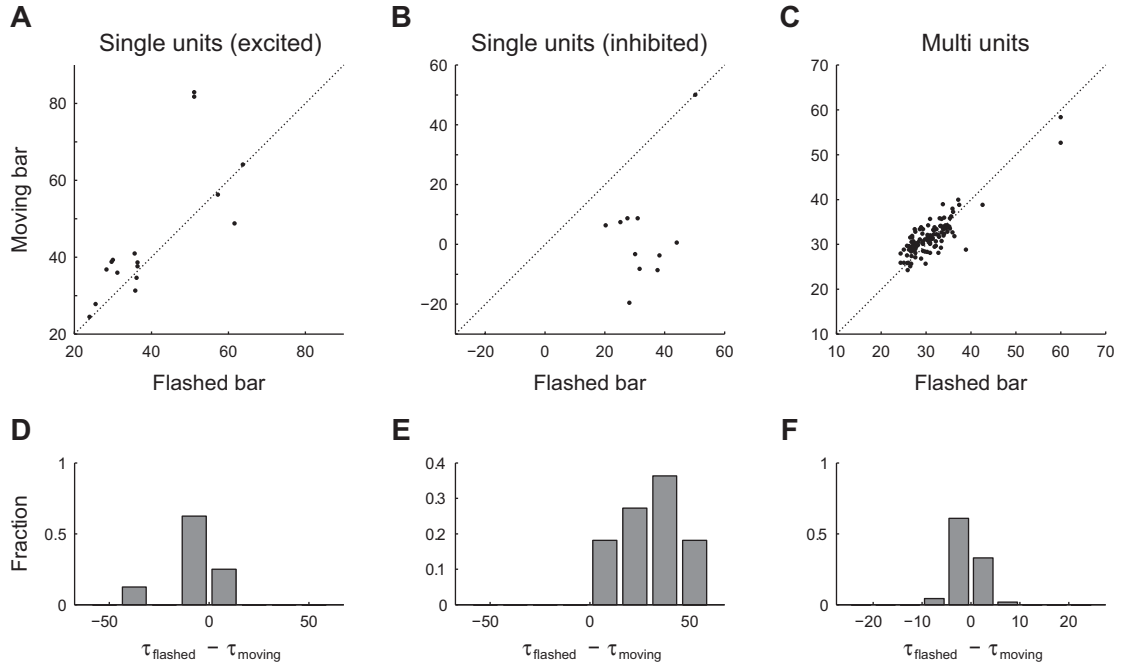


Figure 3.5: Latencies to moving and flashed bars. **A–C:** Scatter plot of latencies to moving bar against latencies to flashed bar. **D–F:** Differences in latencies. While excited single units (A,D) and multi units (C,F) have approximately equal latencies, inhibited single units (B,E) stop responding significantly earlier to moving bars than to flashed bars

ones, we classified single units into inhibited or excited and analyzed the two groups separately.

Fig. 3.5 shows the results of this analysis for all three groups (excited and inhibited single units, multi units). Surprisingly, we find that multi units seem to respond slightly slower to moving bars than to flashed bars (flashed: $31.0 \pm 5.2$ ms [mean $\pm$ s.d.]; moving: $31.7 \pm 4.5$ ms; paired t-test: $p = 0.0035$; $n = 115$). In contrast, for excited single units there is no significant difference between response times to moving and flashed bars (flashed: $39.6 \pm 13.1$ ms; moving: $45.0 \pm 17.6$ ms; paired t-test: $p > 0.05$; $n = 16$) and for inhibited single units, inhibition kicks in significantly earlier for the moving bar than for the flashed bar (flashed: $33.1 \pm 8.7$ ms; moving: $3.5 \pm 17.8$ ms; paired t-test:$p = 0.0001$; $n = 11$). This means that these cells get inhibited around the time the moving stimulus arrives in the center of their receptive field.

However, there is one major confound in the results reported above which needs to be addressed. The stimulus used for the reverse correlation was changing location every second frame, while the moving stimulus moved to its new location every frame. Taking one frame as the discrete time unit and assuming both frames have the same influence on the evoked firing rate, this implies that the estimated receptive field is a convolution of the true receptive field with a two bins wide box kernel. Unfortunately, if the first frame contributes more to the evoked response than the subsequent frame, the "center of mass" of the contribution will be shifted towards the first frame. Thus, the filter kernel (which is centered between the two frames) will introduce a filter delay, leading to an underestimation of the response latency. In the worst case scenario, the response

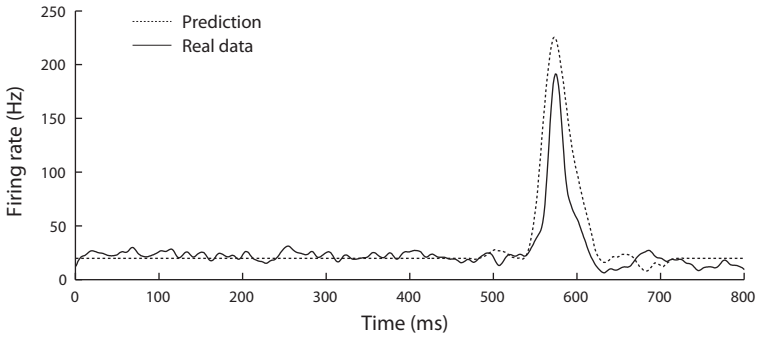


Figure 3.6: Example of spike density function and linear prediction obtained from the space/time receptive field $D(x, \tau)$ (for details on how predictions are computed, see section 2.4.2.1).

is evoked solely by the onset of the flash and the second frame does not contribute at all. In this case, the filter will bias the response latency estimate by half a frame or 5 ms.

From this two conclusions concerning the results presented above can be drawn. On the one hand, because the first frame is very likely to contribute more than the second frame, the fact that responses to moving bars are significantly slower than to flashed bars seems to be an artifact of the way the stimulus is presented. On the other hand, the exact way in which the two frames contribute to the evoked response is unknown (and this might even be different between neurons). Thus, for excited units a clear result about the exact difference in latencies cannot be obtained from the currently available data and further studies are needed. However, what can be concluded from the present data is that (a) inhibited units do show a significantly earlier response onset to moving stimuli than to flashed ones, and (b) if there is a difference in latencies between moving and flashed stimuli for excited units in V1, then this difference is relatively small (< 5 ms) compared to what would be necessary to explain an effect of the magnitude of the FLE.

### 3.2.3 Response latencies predicted by a linear model

Berry et al. (1999) report shorter latencies to moving compared to flashed stimuli in the retina and show that this behavior can be explained on the basis of spatially extended receptive fields and contrast-gain control. It is important to point out that their model makes the opposite prediction about speed dependency of the flash-lag effect from what is experimentally observed (Nijhawan 1994; Krekelberg and Lappe 1999)—a fact that is not acknowledged in their paper. Specifically, if neurons start firing earlier to moving stimuli than to flashed ones because of the moving stimulus entering their receptive field some time before it reaches the center, this predicts that neurons will fire earlier to slowly moving stimuli than to fast moving ones. The magnitude of the flash-lag effect, however, increases as a function of speed (Nijhawan 1994; Krekelberg and Lappe 1999).

Therefore, we next asked whether the results obtained in the previous section could be explained by the spatial extent of receptive fields alone. If this is the case, these data cannot be interpreted as evidence for the differential latencies hypothesis. If the measured responses, however, differ from what is predicted by the cells' receptive fields, this could be an indication for additional (potentially non-linear) mechanisms contributing to the flash-lag illusion.

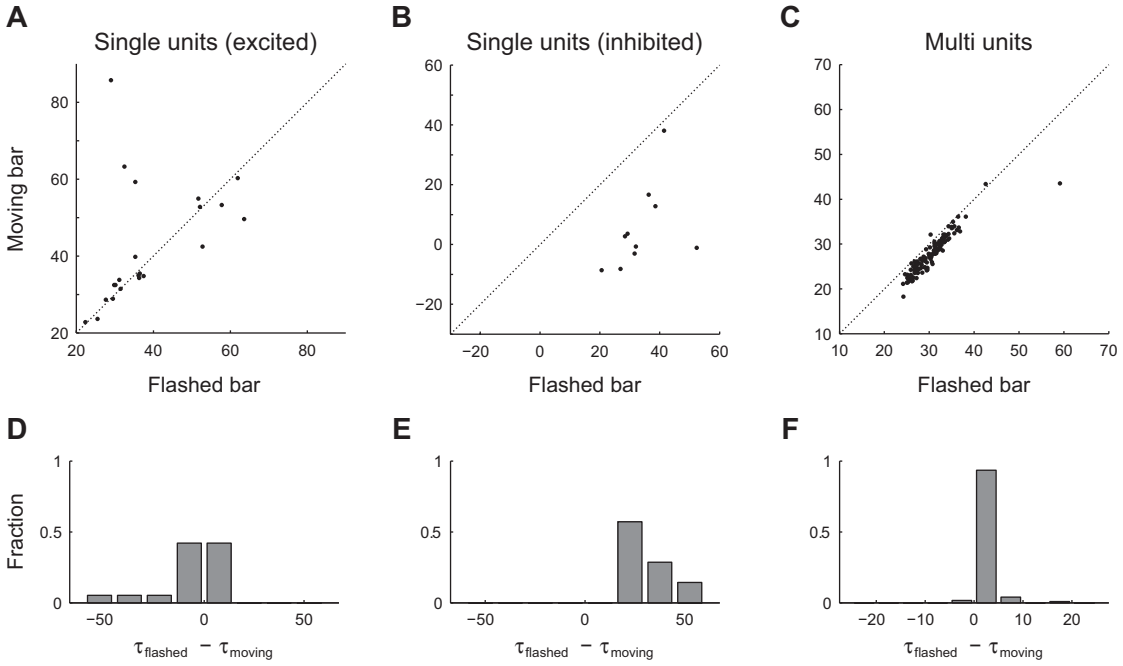


Figure 3.7: Predicted latencies to moving and flashed bars. Plots as in Fig. 3.5.

To address this question, we generated linear predictions of the firing rates in response to the moving bar stimulus using the cells' linear receptive fields (see section 2.4.2.1 for details). Fig. 3.6 shows an example prediction plotted along with the measured spike density function. Note that although it appears slightly shifted in time, it is able to reproduce the general shape of the response relatively well. Given that for most of the cells the shape of response could be predicted well, we repeated the analysis of the previous section for the predicted data.

Fig. 3.7 displays the same plots as Fig. 3.5 but this time it was derived from the linear predictions of the firing rates. Interestingly, the predictions reproduce the measured data relatively well, with the main difference being that for excited units response onsets are predicted to be slightly earlier than observed in the real data. Excited single units are predicted to respond 3.4 ms faster (sign test: $p = 0.005$; $n = 21$) and multi units 4.1 ms faster (sign test: $p = 8.4 \times 10^{-21}$; $n = 124$). In contrast, responses of inhibited single units are predicted very accurately by the linear model (median difference 0.4 ms; sign test: $p = 1$; $n = 10$).

Given that observed latencies are slightly *slower* than predicted by the receptive field structure alone, these results suggest that the receptive fields are indeed biased towards lower latencies, as suspected in the previous section. Additionally, the results are evidence against the differential latencies hypothesis (at least in V1): Latency differences found in our data can be explained entirely on the basis of spatially extended receptive fields, which makes them scale in the opposite direction as the flash-lag illusion: At slower speeds, higher latency differences would be expected, whereas in the flash-lag illusion, the reverse is true.

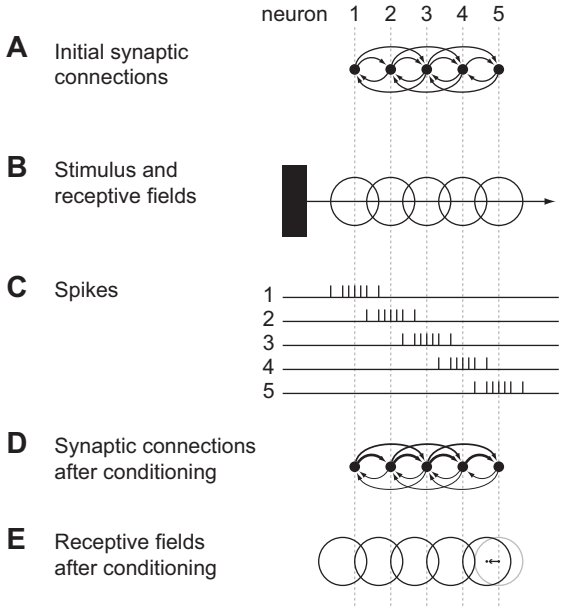


Figure 3.8: Illustration of network with lateral connections being altered by predominantly rightwards moving stimulus. **A:** A small toy network containing five neurons is shown. Each neuron is recurrently connected to each other neuron in the network (for clarity, only connections to the next two neighbors on each side are shown). Initially, all synaptic weights are equal. **B:** The Stimulus (black bar) is moving from left to right across the receptive fields of the five neurons, depicted by circles. **C:** Hypothetical spike trains. Due to their spatially arranged receptive fields, the neurons fire in sequential order. **D:**. Synaptic weights after repeated conditioning with the rightwards moving stimulus. Because neuron 3 is repeatedly activated right before neuron 4, the synapse from neuron 3 to neuron 4 is strengthened. Neuron 3 is also activated before neuron 5 but with larger temporal distance. Thus, the synapse from neuron 3 to neuron 5 is not potentiated as much. Since neurons 1 and 2 are activated before neuron 3, the synapsed to those neurons are weakened depending on the distance of their receptive fields. **E:** After conditioning, synapses coming in from the left side are potentiated whereas synapses from the right side are weakened. This creates an asymmetry in the circuit: If a stimulus appears on the left side of a neuron's receptive field, it will receive strong excitatory input from its neighboring neurons, whereas a stimulus on the right side of the receptive field will not excite the neurons as much. This creates a "predictive receptive field shift" in the direction opposite to the direction of motion.

## 3.3 Stimulus-induced changes in receptive field structure

In the second part of our experiment we asked whether manipulating the statistical properties of the environment could induce changes in the firing characteristics of V1 neurons as predicted by a spike timing dependent plasticity rule (Abbott and Blum 1996; Rao and Sejnowski 2003). The statistical regularity we are systematically modifying in this experiment is the probability of the bar moving leftwards versus it moving rightwards.

If spike timing is indeed crucial for synaptic plasticity and this plasticity operates on time scales on the order of several minutes in awake animals, this should, because

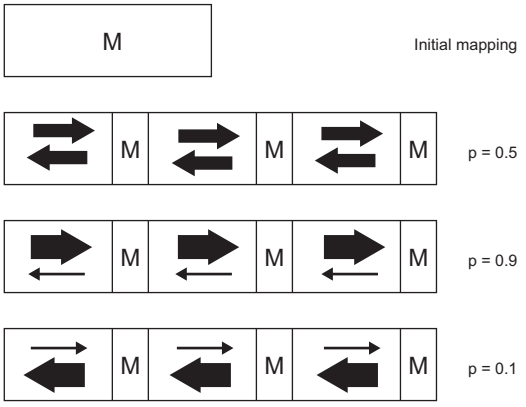


Figure 3.9: Illustration of motion conditioning paradigm. After initial receptive field mapping (first row, depicted by M) three motion conditioning blocks follow. Each block consists of interleaved conditioning and receptive field mapping subblocks. The size of the arrows represents the probability of the bar moving in the respective direction during conditioning trials. M depicts receptive field mapping blocks. Time goes from top to bottom and from left to right.

of lateral connectivity, induce shifts in the neurons' receptive fields. Fig. 3.8 serves to illustrate this point. It shows a small example network of five neurons, ordered by the location of their receptive fields and connected by excitatory lateral connections. If a bar is repetitively moving over the receptive fields from left to right, the neurons in the network will get activated in sequential order (Fig. 3.8 C). Thus, neuron 3 will typically fire spikes slightly after neuron 2 but shortly before neuron 4. Because of the asymmetric spike timing dependent plasticity rule (Fig. 1.3 on page 12), this will strengthen synapses from early to later activated neurons and weaken those from later to earlier neurons (Rao and Sejnowski 2001, 2003; Yao and Dan 2001; Fu et al. 2002). The magnitude of synaptic potentiation and depression will depend on the time interval between the fired spikes and, hence, on the distance of the neurons' receptive fields.

As a consequence of this repetitive stimulation and the resulting synaptic modifications, the neurons' receptive fields should shift to the left, i. e. opposite to the direction of motion. To understand why this is the case, consider neuron 3. If the bar is moving from left to right, neurons 1 and 2 will get activated first and due to their strong excitatory connections, neuron 3 will start firing slightly before the bar reaches its receptive fields. If, however, the bar moves in the opposite direction, neurons 4 and 5, which have only very weak connections to neuron 3 will not cause it to fire earlier, creating an asymmetry that effectively shifts neuron 3's receptive field to the left.

### 3.3.1 Paradigm

The paradigm consisted of an initial receptive field mapping phase of typically ~150 trials. During receptive field mapping, vertically oriented bars were flashed at random locations (see methods for details). After initial receptive field mapping, three blocks with three subblocks each followed. Each subblock consisted of 85 motion conditioning trials and 15–50 receptive field mapping trials. In motion conditioning trials, a vertically oriented bar was moving across the visual field either in leftwards or in rightwards

direction. The probability of motion direction was varied between blocks. During the first block, the probability of the bar moving rightwards was 50%; in the second block it was 90%; in the last block it was 10%. The sequence of a full experiment is depicted in Fig. 3.9.

### 3.3.2 Data sample

The data set used for this analysis consisted of 6 recording sessions collected over the course of two weeks. 55 single units and 131 multi units were recorded in this time. Some of the units were discarded because they did not fire a sufficient number of spikes to obtain reliable receptive fields or the receptive fields did not show clear structure (i. e. were not fit well by a Gaussian). After discarding those units, the data sample contained 37 single units and 126 multi units.

### 3.3.3 Effect of motion conditioning on receptive field location

To quantify the dependence of receptive field structure on the prior probability of motion direction, we first analyzed the receptive field center locations as a function of the prior probability of motion direction. According to our hypothesis, receptive fields should shift to the left in the second block ($p = 0.9$), because the bar is moving from left to right in 90% of the trials. In contrast, in the third block ($p = 0.1$) they should shift to the right.

Centers were extracted by taking the mean of a one-dimensional Gaussian fitted to the spatial receptive field component (see methods for details). To account for different absolute receptive field locations between cells, the mean was subtracted for each cell individually. Fig. 3.10 shows the distributions of receptive field centers for each of the three different priors. For the single units, the median receptive field centers are $0.73'$, $-1.45'$, and $0.32'$ for the priors $p = 0.5$, $p = 0.9$, and $p = 0.1$, respectively. Although these values show a trend in the direction we expected, they were not significantly different (non-parametric Friedman test: $p = 0.097$; $\chi^2 = 4.67$; df $= 62$). For the multi units, receptive field centers were indistinguishable between priors ($-0.03'$, $-0.03'$, and $0.06'$; Friedman test: $p = 0.90$; $\chi^2 = 0.20$; df $= 377$). In summary, these data do not support the initial hypothesis that a stimulus which moves very predictably in one of two directions will induce a shift in receptive fields of V1 neurons.

### 3.3.4 Effect of motion conditioning on receptive field skewness

Since receptive field centers did not change as a function of the prior, we next asked whether receptive fields might instead become skewed. Mehta et al. (2000) report hippocampal place fields to become skewed opposite to the direction of movement as rats became familiar with their environment. They show that this skew is predicted by a simple feedforward network model with a temporally asymmetric STDP rule where the amount of synaptic potentiation is larger than the amount of depression. Because of this asymmetry in potentiation and depression, the first spikes a neuron is firing when the moving bar enters its receptive fields are expected to undergo a larger predictive shift than the later spikes (Mehta et al. 2000). Therefore, we tested whether the predominantly rightwards (leftwards) moving bar in our experiment would introduce a leftwards (rightwards) skew in the measured receptive fields.

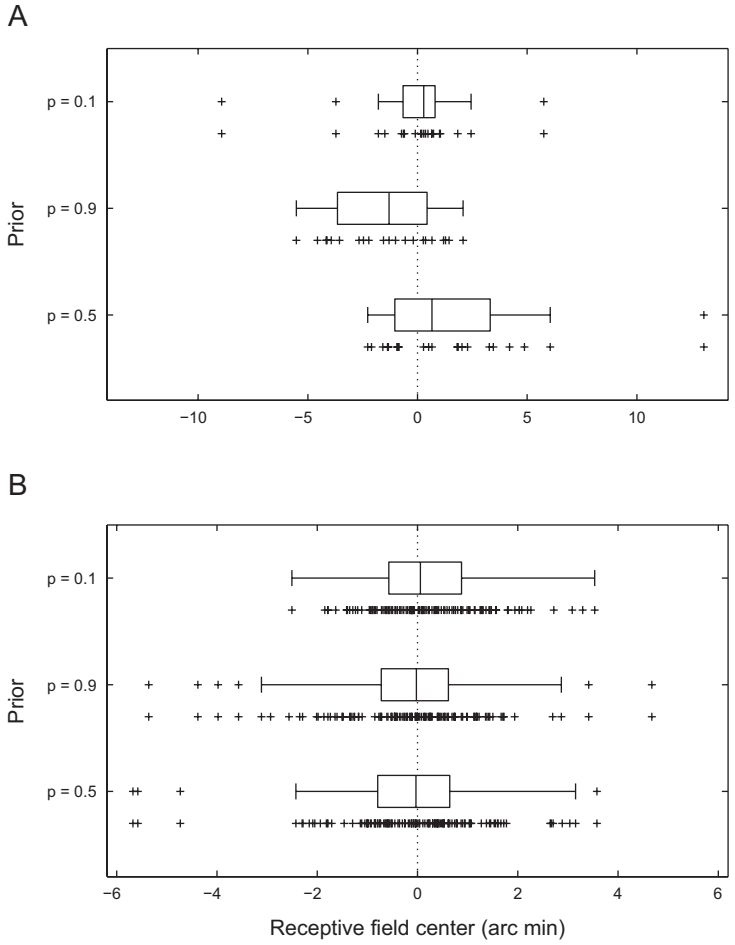


Figure 3.10: Distribution of receptive field (RF) center locations as a function of prior probability of the bar moving rightwards. Because data deviate quite strongly from normal distribution, box plots and the actual data points are shown and non-parametric statistical tests are used. Center lines in boxed represent median, boxes are quartiles, and whiskers denote the range. Data points further away than 2.5 times the interquartile range are considered outliers. Raw data values are plotted below the box plots. **A:** Single units. No effect of prior (Friedman's test: $p = 0.097$). **B:** Multi units. No effect of prior (Friedman's test: $p = 0.90$).

Fig. 3.11 shows that we did not observe such a predictive skew in either single or multi unit receptive fields. For single units, skewness is statistically indistinguishable between priors (median skewness $\gamma_{0.5} = -0.004$, $\gamma_{0.9} = 0.001$, and $\gamma_{0.1} = 0.013$; Friedman's test: $p = 0.28$, $\chi^2 = 2.57$, df $= 62$). Surprisingly, multi unit receptive fields were initially slightly skewed to the left and became more symmetric in subsequent blocks ($\gamma_{0.5} = -0.13$, $\gamma_{0.9} = -0.048$, and $\gamma_{0.1} = -0.051$; Friedman's test:$p = 0.014$, $\chi^2 = 8.59$, df $= 377$). Post-hoc testing revealed a significant difference between $\gamma_{0.5}$ and $\gamma_{0.9}$ but not between any other pair. This result is relatively difficult to interpret. If there was a real effect of motion conditioning, one would expect it to be observable not only between the $p = 0.5$ and $p = 0.9$ block but also between $p = 0.9$ and $p = 0.1$. Additionally, the observed difference is in the opposite direction of what has previously been reported

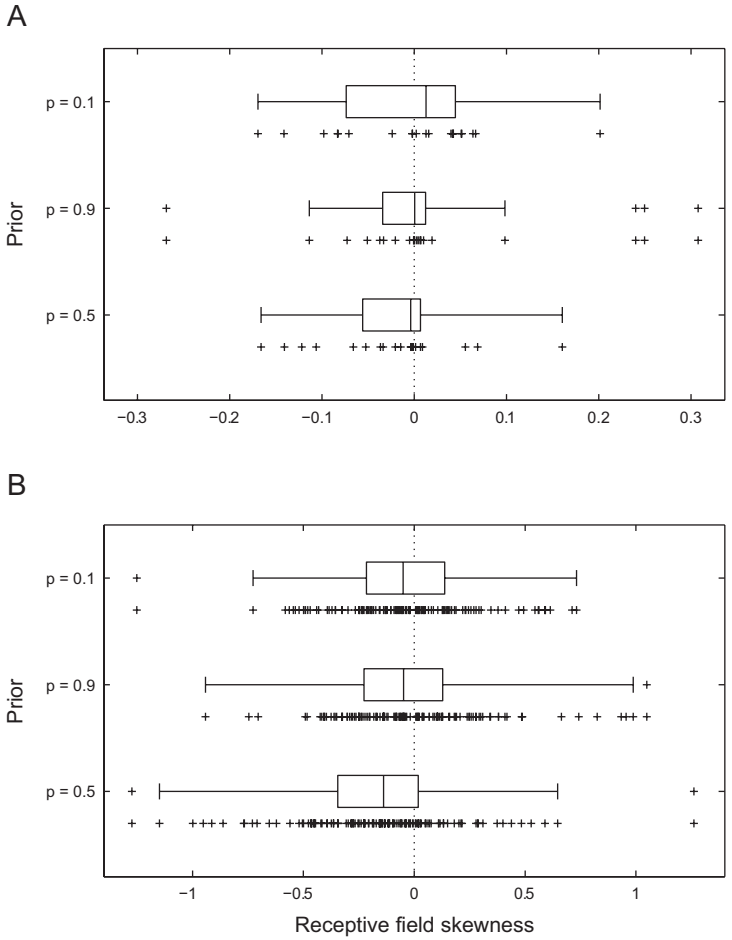


Figure 3.11: Distribution of receptive field skewness. Plots as in Fig. 3.10. **A:** Single units. No effect of prior (Friedman's test: $p = 0.28$). **B:** Multi units. Effect of prior (Friedman's test: $p = 0.014$). Skewness is different between $p = 0.5$ and $p = 0.9$ according to post-hoc testing. See text for possible explanation.

and was predicted by models (Mehta et al. 2000). We therefore conclude that this result is most likely due to natural statistical fluctuations.

## 3.4 Effects of varying the prior on response latencies

In the previous section we analyzed whether receptive fields are shifted by motion conditioning. This analysis assumed that repeated processing of a stimulus predominantly moving in one of two directions affects the way neurons fire in response to statically flashed bars. Although this was predicted by models (Abbott and Blum 1996; Rao and Sejnowski 2003) and has been shown to occur in anesthetized cats (Fu et al. 2002; Yao and Dan 2001), the negative result obtained above hints at the possibility that in the awake animal such effects might be more stimulus-specific. We therefore consider in this section the neural responses to the moving bars under the three different priors of motion direction.

We computed spike density functions (SDF) for each prior and direction of motion in 1 ms bins. Those were further smoothed by a Gaussian kernel with a standard deviation

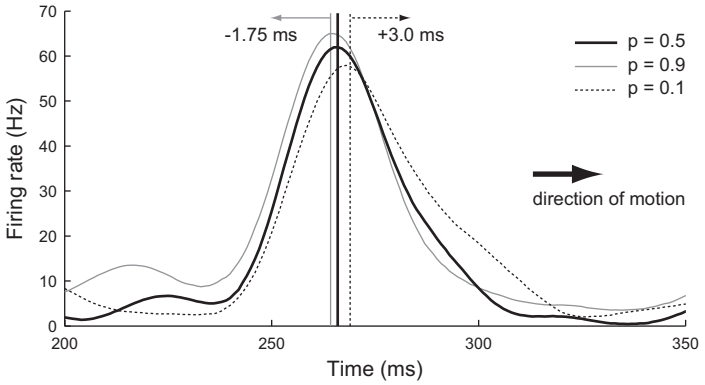


Figure 3.12: Example of a spike density function shifting opposite to the direction of motion.

of 8 ms. Upon visual inspection, we observed that in the $p = 0.9$ and $p = 0.1$ blocks, some neurons' responses to the direction of motion presented in 90% of the trials shifted opposite to the direction of motion, whereas the opposite was true for the responses to the 10% direction. Fig. 3.12 shows such an example.

In order to quantify this effect on the population level, we determined the position of each SDF in time. This was done by shifting the SDF relative to a reference until the mean squared error between the two was minimized. As the reference, the SDF obtained from the $p = 0.5$ block was used. Only neurons for which the correlation between the optimally matched spike density functions was at least $r > 0.95$ were included in the analysis. To avoid a potential bias due to the different number of trials that go into the computation of SDFs, we use the same number of trials in all six conditions (2 directions $\times$ 3 priors). The minimum number of trials used was between 20 and 24, depending on the experimental session. If there were more trials available for a condition, we selected a uniformly sampled subset of trials.

Fig. 3.13 A, C, and E show the amount by which responses of excited single units, inhibited single units, and multi units, respectively, are shifted in each block. The SDF obtained from the entire $p = 0.5$ block was used as the reference; thus, the shifts for this block are expected to be close to zero and can serve as an estimate of how reliable the results are.

For they gave the most reliable and significant results, we start by reporting our observations about the multi units. A two factor analysis of variance (ANOVA) with factors direction and prior indicated both a significant main effect of prior ($p = 9.7 \times 10^{-10}$) and direction ($p = 0.042$), as well as a significant interaction ($p = 3.0 \times 10^{-4}$). While the main effect of prior could be caused by long term adaptation or general fatigue of the monkey, the main effect of direction is most likely due to the unbalanced experimental design (always the $p = 0.9$ block leads the $p = 0.1$ block). Although it remains to be investigated what caused the main effects of prior and direction, the term we are particularly interested in is the interaction between prior and direction. The same data as in Fig. 3.13 E, corrected for main effects, is plotted in Fig. 3.13 F. It can be seen that—in addition to what is explained by main effects—between the $p = 0.5$ and $p = 0.9$ blocks processing of rightwards moving bars gets faster while processing of leftwards moving bars gets slower, whereas between the $p = 0.9$ and $p = 0.1$ blocks the opposite is

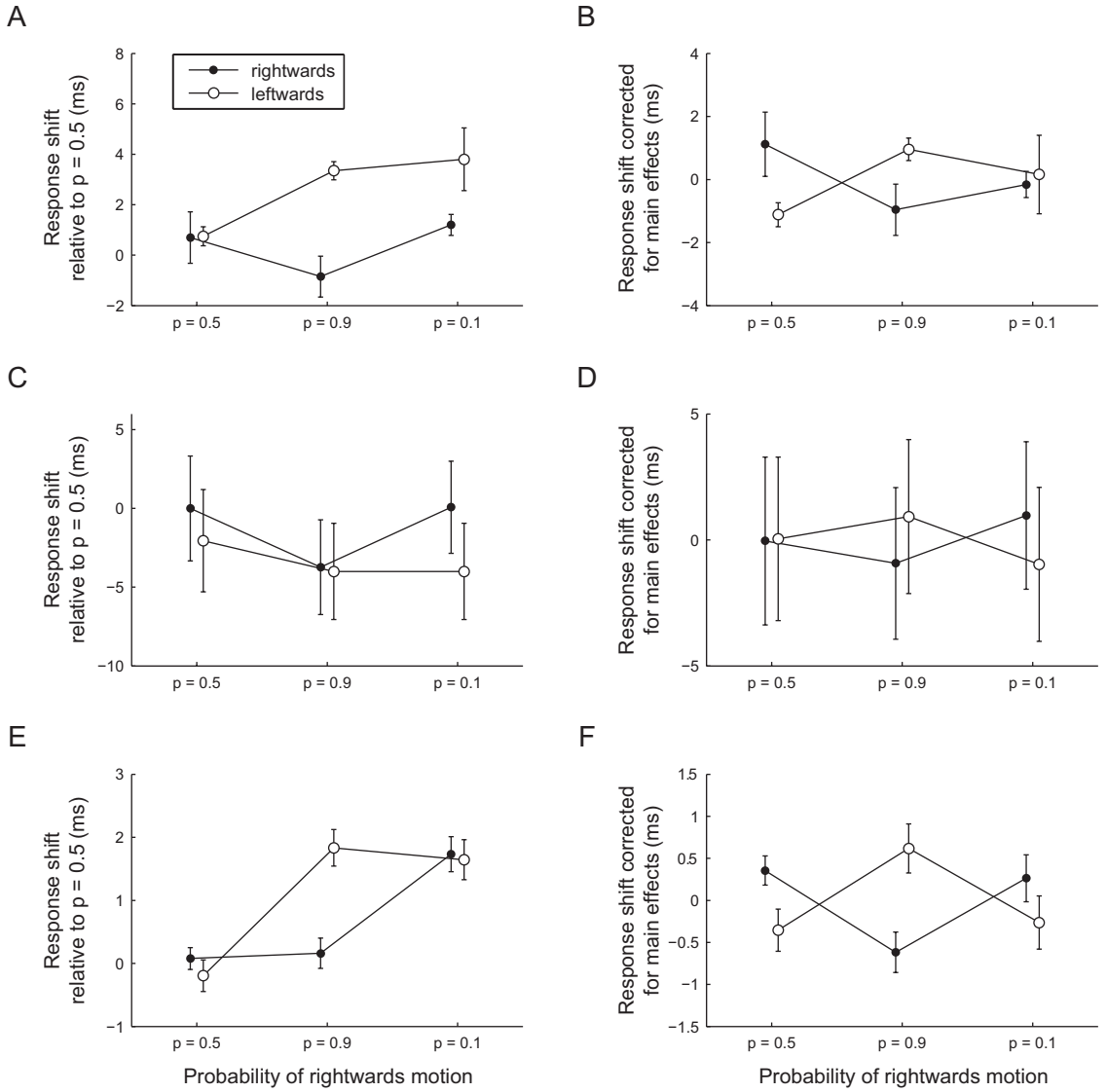


Figure 3.13: Response shifts relatative to $p = 0.5$ block.

true. This interaction is both highly significant ($p = 3.0 \times 10^{-4}$) and predicted by our hypothesis.

In the single units, the picture is less clear, potentially attributable to the smaller sample size. For excited single units a general picture similar to the multi units emerges, although there is no significant main effect of prior ($p = 0.087$). There is, as for the multi units, a significant main effect of direction ($p = 0.0016$) and a significant interaction ($p = 0.044$), which looks comparable (Fig. 3.13 B) to the interaction observed in the multi units.

Inhibited single units, however, do not show any significant effects (prior: $p = 0.65$; direction: $p = 0.40$; interaction:$p = 0.83$). This is not particularly unexpected, since these neurons do not fire spikes in response to the stimulus.

# 4 Discussion

## 4.1 The flash-lag effect

### 4.1.1 Psychophysical results

In the first part of this study, we have demonstrated the flash-lag effect in the continuous motion condition in human observers. We found a significant difference in effect magnitude (point of subjective equality; PSE) between subjects. While this could be due to natural variations between subjects, it could also have been caused by uncontrolled eye or head movements or slightly different eye/monitor distances. Additionally, the subjects included the author and none of them was completely naïve to the illusion. However, the flash-lag effect is a remarkably strong illusion and still very well measurable in all subjects. In summary, the human study showed that our stimulus causes a strong flash-lag effect with a PSE shifted by 0.5°–1° in the direction of the moving object, replicating previous results with linear motion (Whitney and Murakami 1998; Whitney et al. 2000).

With our monkey psychophysics study, we have shown for the first time that monkeys also perceive the flash-lag illusion. This was achieved by training a monkey to report his percept using a special rewarding scheme which ensures that he is not trained the response by reward but truly reports his percept. The magnitude of the monkey flash-lag effect appears comparable to the one in humans. Naturally, to obtain a more reliable comparison between species, more thoroughly controlled human psychophysics are necessary and parameters have to be exactly matched. In this case, however, some parameters had to be adjusted for the monkey after the human pilot study.

Kanai et al. (2004) show contrast dependence of the flash-lag effect magnitude in humans in the flash-termination condition. In our monkey data, where we tested two relatively high contrasts, there is no significant difference between conditions. This is in agreement with Kanai et al. (2004) as they only find an increase in PSE shift for extremely low contrast (Michelson contrast 0.043) but not for contrasts on the order of the values we tested (Michelson contrast 0.30 and 0.43). Unfortunately, low contrast is hard to test in the monkey since performance goes down drastically, making it very hard to fit psychometric functions.

In contrast, a study of speed dependence of the monkey FLE should be relatively straightforward and is indeed being carried out at the moment. Whether replication of various other human FLE results such as FIC and FTC are possible in the monkey needs to be investigated. One major problem is that it is unclear if monkeys generalize the learned task (continuous motion condition) to FIC and FTC or if they need to be explicitly trained on each variant.

### 4.1.2 Differential latencies?

Jancke et al. (2004) find differential latencies for moving versus flashed bars in anesthetized cats' area 17. We cannot confirm these findings in awake monkeys on the basis

of our available data. First, for excited single and multi units the difference in latencies is certainly less than 5 ms and can therefore not account for the FLE. Although inhibited single units stop firing significantly earlier for moving versus flashed bars, this is expected due to their spatially very extended receptive fields. Berry et al. (1999) show that in the retina, the spatial extent of RFs can fully account for observed anticipatory firing of ganglion cells. While they claim this to be a first correlate of the FLE as early as in the retina, it needs to be pointed out that their model makes the exact opposite prediction about the speed dependence of the FLE to what is experimentally observed (Nijhawan 1994; Krekelberg and Lappe 1999). As suggested by Krekelberg and Lappe (2001), their model could be modified to fit the psychophysical data if it is assumed that neurons' speed preferences correlate with their receptive field sizes. This would at higher speeds preferentially activate neurons with larger receptive fields and therefore produce a larger predictive shift than at lower speeds, consistent with psychophysical data. As neurons' receptive field sizes increase with eccentricity, this model would predict and increase of FLE magnitude with increasing eccentricity of the moving bar—a dependence which has been found experimentally (Kanai et al. 2004). However, we are not aware of any experimental evidence for a correlation between receptive field size and speed preference in early visual neurons.

There is one confound to our estimated latencies: It's not entirely clear whether the receptive fields derived from responses to the white noise flashing bars stimulus can be equated to the response to a single flash. Although receptive fields are defined as the first Wiener kernel and therefore the impulse response to a single flash, they are derived from a stimulus that is constantly changing and very dynamic, whereas a single flash is a very different context. Thus, neurons' firing could be different (slower) to a single flash compared to when they are generally excited by a constantly changing stimulus such as the white noise one used for RF mapping.

### 4.1.3 Future directions

This leads us to some suggestions for future work. First, a more thorough comparison between monkey and human flash-lag effect could help to better link neurophysiological results to human perception. To this end, we plan to systematically vary speed, contrast, and eccentricity in further studies in both humans and monkeys. Second, responses to both moving and single flashed bars need to be investigated experimentally in awake primates in order to directly test the differential latencies hypothesis. Since there remains some doubt about the validity of studying moving and flashed bars in isolation, we plan to run electrophysiology experiments with the full FLE stimulus in the future. Additionally, to test whether the model proposed by Berry et al. (1999) can be modified to account for the FLE, a possible dependence between receptive field size and speed tuning needs to be investigated.

Finally, and on a more general note, to unravel the role of different areas along the visual processing stream, it is of major importance to study responses of different visual areas under the exact same stimulus. This will potentially enable researchers to better understand information processing and transformations being performed between the retina, LGN, V1, and extrastriate areas.

## 4.2 Spike timing dependent plasticity

### 4.2.1 Effects of motion conditioning

The hypothesis of STDP-based receptive field shifts had to be rejected in section 3.3. Here we discuss some potential explanations for this negative result. Since synaptic modification has been shown to depend on the temporal order of pre- and post-synaptic action potentials (Markram et al. 1997; Zhang et al. 1998; Feldman 2000; Froemke and Dan 2002), one expects the effect magnitude to be dependent both on the total number of spikes fired during the experiment and on the reliability of their temporal sequence. While in-vitro studies find significant changes in synaptic strength with as little as 60–80 pre/postsynaptic spike pairings (Froemke and Dan 2002), in order to induce detectable changes in adult cats in-vivo much more intensive stimulation was necessary (Yao and Dan 2001; Fu et al. 2002; Yao et al. 2004). For instance, 1200 pairs of asymmetric conditioning lead to a ~1.5–2% shift in receptive fields (Fu et al. 2002); 1600 pairs of orientation conditioning resulted in a 2.3° shift in orientation tuning curves. Given that our conditioning paradigm had only ~150 trials per conditioning block and the monkey was allowed to move his eyes between trials, any resulting effect is expected to be small. Thus, it is possible that with our paradigm and the limited amount of data available, we were not able to detect a possible effect. However, if this is indeed the case, the question arises how relevant STDP is in realistic scenarios. Under natural viewing conditions, one is never looking exclusively at 1000 repetitions of the same stimulus sequence

Additionally, the study of receptive field locations after asymmetric conditioning has a serious drawback. One desirable property of a predictive code is certainly a minimal rate of wrong predictions. Therefore, predictive shifts due to conditioning should—if possible—be as specific to the conditioned stimulus as possible. Otherwise it would result in mislocalization for most stimuli other than the conditioned one. As the stimulus used for receptive field mapping is very different from the conditioning stimulus, if effects were stimulus-specific, one would not expect to observe changes in receptive field structure. Thus, it is conceivable that the shifts reported in the above studies (Fu et al. 2002; Yao and Dan 2001; Yao et al. 2004) indeed represent relatively drastic modifications in the neural circuit leading to perceptual illusions, as demonstrated by the psychophysical results.

Indeed, our data suggest that modifications might be stimulus-specific. Analysis of responses to moving bars under three different priors of motion direction reveals significant interactions between direction of motion and prior for both single and multi units. In fact, the interaction effect size of ~1 ms corresponds to the bar moving ~3.5% of the average multi unit receptive field size and, hence, is relatively stronger to what other studies (Fu et al. 2002) reported for receptive field shifts.

However, care has to be taken when interpreting these results. Increasing response latencies during the course of the experiment were found (ANOVA; main effect of prior) as well as—relative to the $p = 0.5$ block—higher latencies to leftwards versus rightwards motion (ANOVA; main effect of direction of motion). Only after correcting for those main effects and considering interaction terms, the data showed the predicted structure. Although a statistically sound analysis, it is unclear to what extend it is confounded by these uncontrolled factors.

Another confound is that a simple explanation for the observed interactions would be a bias in the exact location the monkey is fixating at due to the different priors of motion direction. Although we took great care in recording eye movements, the effective resolution is too low to exclude this possibility based on eye movement data alone. However, because of the main effect of direction of motion revealed by the ANOVA, we can rule out this possibility. A bias in fixation location is not consistent with this effect. Since in the first block, both directions are equally likely to occur, there should not be any bias. In the second block, however, the bias should be in one direction because of rightwards motion occurring more often while in the last block, it should be in the other direction because of leftwards motion being more likely. Taken together, the average bias should be zero. In contrast, if response times are modified due to an STDP-based mechanism, this predicts to first facilitate rightwards motion in the $p = 0.9$ block and subsequently reduce the effect in the $p = 0.1$ block. Overall, this would on average predict faster responses to the rightwards moving bar because of the unbalanced experimental design—which is what is observed.

### 4.2.2 Future directions

One major issue in the present study was the limited amount of data available to study relatively small changes in neural firing properties. It would thus be helpful to record additional data in order to be able to better quantify the observed effects. Furthermore, we are currently investigating how the stimulus can be optimized to increase the effect size. One approach would be to change the prior probabilities in the conditioning blocks to less extreme values (such as, say, $p = 0.8$ and $p = 0.2$) as this would increase the available data for the rarely presented stimulus significantly.

A second issue to be investigated is the source of the increasing response latencies we observed. In combination with a balanced experimental design (randomized order of direction conditioning blocks) this would potentially lead to clearer results in the analysis carried out in section 3.4.

A possibility not investigated in this thesis is that the visual system is computing some kind of a "local prior" of motion direction (i. e. percentage of trials rightwards out of the last 20) and adapts relatively fast to such changes. A way to investigate this scenario could be to smoothly vary the probability of rightwards motion during the course of the experiment and correlate the observed response latencies with the fraction of rightwards trials out of the last $n$ trials.

Finally, the fact that the observed changes are stimulus-specific could be a sign of a mechanism by which neural circuits can learn certain regularities without distorting previously learned relationships. This could have important implications for perceptual learning since one aspect of perceptual learning is that it is also often specific to the learned stimulus configuration and experimental context. For instance, the behavioral improvement during training on a three-line bisection task does not transfer to a vernier acuity task. This sort of specificity may pose problems to perceptual learning models based solely on changes in the basic tuning functions of neurons in V1, since one may think that changes in the basic tuning of functions of neurons should lead to general effects on sensory processing. To circumvent this problem it has been suggested (e. g. Tsodyks and Gilbert 2004) that this content-specific learning may be explained by task-dependent modulation of lateral interactions in V1.

# References


Abbott, L. F. and Blum, K. I. (1996). Functional significance of long-term potentiation for sequence learning and prediction. *Cereb. Cortex*, 6:406–416.

Berry, M. J., Brivanlou, I. H., Jordan, T. A., and Meister, M. (1999). Anticipation of moving stimuli by the retina. *Nature*, 398:334–338.

Bi, G. Q. and Poo, M. M. (1998). Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *Journal of Neuroscience*, 18(24):10464–72. United states.

Bliss, T. V. P. and Lomo, T. (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *J Physiol*, 232(2):331–356.

Boch, R. and Fischer, B. (1986). Further observations on the occurrence of express-saccades in the monkey. *Experimental Brain Research*, 63(3):487–494.

Bolz, J. and Gilbert, C. D. (1986). Generation of end-inhibition in the visual cortex via interlaminar connections. *Nature*, 320:362–365.

Brainard, D. H. (1997). The psychophysics toolbox. *Spat Vis*, 10(4):433–436.

Dayan, P. and Abbott, L. F. (2001). *Theoretical Neuroscience*. MIT Press, Cambridge, MA.

Digital Signal Processing Committe (1979). *Programs for Digital Signal Processing*. IEEE Press, Piscataway, NJ, USA.

Eagleman, D. M. and Sejnowski, T. J. (2000). Motion integration and postdiction in visual awareness. *Science*, 287:2036–2038.

Eagleman, D. M. and Sejnowski, T. J. (2007). Motion signals bias localization judgments: a unified explanation for the flash-lag, flash-drag, flash-jump, and Frohlich illusions. *Journal of vision*, 7(4):1–12.

Ecker, A. S., Berens, P., Keliris, G. A., Logothetis, N. L., and Tolias, A. S. (2007a). A data management system for electrophysiological data analysis. In *Proceedings of the 7th Meeting of the German Neuroscience Society*, volume 7, Göttingen.

Ecker, A. S., Berens, P., and Tolias, A. S. (2007b). A data management system for electrophysiological data analysis. Technical report, Karl Steinbuch Stipendium, MFG Stiftung Baden-Württemberg.

Feldman, D. E. (2000). Timing-based LTP and LTD at vertical inputs to layer II/III pyramidal cells in rat barrel cortex. *Neuron*, 27(1):45–56.

Froemke, R. C. and Dan, Y. (2002). Spike-timing-dependent synaptic modification induced by natural spike trains. *Nature*, 416(6879):433–8.

Fu, Y.-X., Djupsund, K., Gao, H., Hayden, B., Shen, K., and Dan, Y. (2002). Temporal specificity in the cortical plasticity of visual space representation. *Science*, 296:1999–2003.

Fu, Y.-X., Shen, Y., Gao, H., and Dan, Y. (2004). Asymmetry in visual cortical circuits underlying motion-induced perceptual mislocalization. *J. Neurosci.*, 24:2165–2171.

Gattass, R., Gross, C. G., and Sandell, J. H. (1981). Visual topography of V2 in the macaque. *The Journal of Comparative Neurology*, 201(4):519–539.

Hazelhoff, F. F. and Wiersma, H. (1924). Die Wahrnehmungszeit. *Zeitschrift für Psychologie*, 96:171–188.

Hebb, D. O. (1949). *The Organization of Behavior*. John Wiley & Sons Inc, New York.

Helmholtz, H. v. (1896). *Handbuch der Physiologischen Optik*. Leopold Voss, Hamburg and Leipzig, 2 edition.

Hubel, D. H. and Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *J Physiol*, 195:215–243.

Ito, M. (1986). Long-term depression as a memory process in the cerebellum. *Neurosci Res*, 3(6):531–539.

Ito, M. (1989). Long-term depression. *Annual Review of Neuroscience*, 12(1):85–102.

Jancke, D., Erlhagen, W., Schöner, G., and Dinse, H. R. (2004). Shorter latencies for motion trajectories than for flashes in population responses of cat primary visual cortex. *Journal of Physiology*, 556(3):971–82.

Judge, S. J., Wurtz, R. H., and Richmond, B. J. (1980). Vision during saccadic eye movements. I. Visual interactions in striate cortex. *J Neurophysiol*, 43(4):1133–55.

Kanai, R., Sheth, B. R., and Shimojo, S. (2004). Stopping the motion and sleuthing the flash-lag effect: spatial uncertainty is the key to perceptual mislocalization. *Vision Research*, 44:2605–2619.

Krekelberg, B. and Lappe, M. (1999). Temporal recruitment along the trajectory of moving objects and the perception of position. *Vision Research*, 16:2669–2679.

Krekelberg, B. and Lappe, M. (2001). Neuronal latencies and the position of moving objects. *Trends in Neurosciences*, 24:335–339.

Krekelberg, B., Lappe, M., Whitney, D., Cavanagh, P., Eagleman, D. M., and Sejnowski, T. J. (2000). The position of moving objects. *Science*, 289:1107a.

Lappe, M. and Krekelberg, B. (1998). The position of moving objects. *Perception*, 27:1437–49.

Levy, W. B. and Steward, O. (1983). Temporal contiguity requirements for long-term associative potentiation/depression in the hippocampus. *Neuroscience*, 8(4):791–797.

Markram, H., Lubke, J., Frotscher, M., and Sakmann, B. (1997). Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science*, 275:213–215.

Mehta, M. R., Barnes, C. A., and McNaughton, B. L. (1997). Experience-dependent, asymmetric expansion of hippocampal place fields. *Proceedings of the National Academy of Sciences of the United States of America*, 94(16):8918–21.

Mehta, M. R., Quirk, M. C., and Wilson, M. A. (2000). Experience-dependent asymmetric shape of hippocampal receptive fields. *Neuron*, 25(3):707–15.

Michelson, A. A. (1927). *Studies in Optics*. The University of Chicago Press.

Nijhawan, R. (1994). Motion extrapolation in catching. *Nature*, 370:256–257.

Nijhawan, R. (1997). Visual decomposition of colour through motion extrapolation. *Nature*, 386:66–69.

Nijhawan, R. (2002). Neural delays, visual motion and the flash-lag effect. *Trends in Cognitive Sciences*, 6:387–393.

Patel, S. S., Ogmen, H., Bedell, H. E., and Sampath, V. (2000). Flash-lag effect: differential latency, not postdiction. *Science*, 290:1051.

Purushothaman, G., Patel, S. S., Bedell, H. E., and Ogmen, H. (1998). Moving ahead through differential visual latency. *Nature*, 396(6710):424.

Rao, R. P. and Sejnowski, T. J. (2001). Spike-timing-dependent hebbian plasticity as temporal difference learning. *Neural Computation*, 13(10):2221–37.

Rao, R. P. N. and Sejnowski, T. J. (2003). Self-organizing neural systems based on predictive learning. *Philosophical Transactions of the Royal Society London, Series a (Mathematical, Physical & Engineering Sciences)*, 361(1807):1149–75.

Sheth, B. R., Nijhawan, R., and Shimojo, S. (2000). Changing objects lead briefly flashed ones. *Nat Neurosci*, 3(5):489–495.

Stent, G. S. (1973). A physiological mechanism for Hebb's postulate of learning. *Proc Natl Acad Sci U S A*, 70(4):997–1001.

Thakur, P. H., Lu, H., Hsiao, S. S., and Johnson, K. O. (2007). Automated optimal detection and classification of neural action potentials in extra-cellular recordings. *Journal of Neuroscience Methods*, 162(1-2):364–376.

Tolias, A. S., Ecker, A. S., Siapas, A. G., Hoenselaar, A., Keliris, G. A., and Logothetis, N. K. (2007). Recording chronically from the same neurons in awake, behaving primates. *J Neurophysiol*, 98(6):3780–3790.

Tsodyks, M. and Gilbert, C. (2004). Neural networks and perceptual learning. *Nature*, 431(7010):775–781.

Ueda, N., Nakano, R., Ghahramani, Z., and Hinton, G. E. (2000). SMEM algorithm for mixture models. *Neural Comp.*, 12(9):2109–2128.

Watanabe, K., Nijhawan, R., Khurana, B., Shimojo, S., Watanabe, K., Nijhawan, R., Khurana, B., and Shimojo, S. (2001). Perceptual organization of moving stimuli modulates the flash-lag effect. *Journal of Experimental Psychology: Human Perception & Performance*, 27(4):879–94.

Watanabe, K., Nijhawan, R., and Shimojo, S. (2002). Shifts in perceived position of flashed stimuli by illusory object motion. *Vision Research*, 42(24):2645–2650.

Watanabe, K., Sato, T. R., and Shimojo, S. (2003). Perceived shifts of flashed stimuli by visible and invisible object motion. *Perception*, 32(5):545–59.

Whitney, D. and Murakami, I. (1998). Latency difference, not spatial extrapolation. *Nat Neurosci*, 1:656–657.

Whitney, D., Murakami, I., and Cavanagh, P. (2000). Illusory spatial offset of a flash relative to a moving stimulus is caused by differential latencies for moving and flashed stimuli. *Vision Research*, 40:137–149.

Wichmann, F. A. and Hill, N. J. (2001a). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & psychophysics*, 63:1293–313.

Wichmann, F. A. and Hill, N. J. (2001b). The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Perception & psychophysics*, 63:1314–29.

Yao, H. and Dan, Y. (2001). Stimulus timing-dependent plasticity in cortical processing of orientation. *Neuron*, 32(2):315–23.

Yao, H., Shen, Y., and Dan, Y. (2004). Intracortical mechanism of stimulus-timing-dependent plasticity in visual cortical orientation tuning. *PNAS*, 101:5081–5086.

Zhang, L. I., Tao, H. W., Holt, C. E., Harris, W. A., and Poo, M. (1998). A critical window for cooperation and competition among developing retinotectal synapses. *Nature*, 395(6697):37–44.

# Acknowledgments

First and foremost I would like to thank my advisor Andreas Tolias for his great support and encouragement over the years and during the work on this thesis. He was always extremely dedicated to help and I don't want to know how often I woke him up or bothered him at 3am.

For this thesis, I worked in close collaboration with Philipp Berens. I am very grateful for excellent team work, without which I am sure I could not have accomplished major parts of the work.

Furthermore, I want to thank the entire Tolias lab who contributed substantially to the successful completion of this thesis. Specifically, Manivannan Subramaniyan for continuing the electrophysiology recordings after we had left Houston, James Cotton for his dedication in getting recording hardware and state system running in time, Andreas Hoenselaar for helping out with monkey training and collecting psychophysical data, and Dennis Murray for his amazing talent in keeping a bunch of chaotic scientists organized.

Last but not least, I would like to thank my girlfriend Anne for being so patient with me during the last few weeks!

# Eidesstattliche Erklärung

Die vorliegende Arbeit wurde im Rahmen eines kollaborativen Forschungsprojektes zwischen September 2007 und März 2008 am Baylor College of Medicine (Houston, TX) und dem Max-Planck-Institut für biologische Kybernetik (Tübingen) durchgeführt. Da die Durchführung derartiger Versuche äußerst zeitaufwändig ist, wurde das dieser Arbeit zugrundeliegende Projekt in enger Zusammenarbeit mit einem weiteren Diplomanden (Philipp Berens, betreut von Prof. Dr. Hauck) durchgeführt, für dessen Arbeit die gewonnen Daten ebenso die Grundlage bilden.

Mit Ausnahme der Kapitel 2.1, und 2.4.1, die im Wesentlichen auf früheren Publikationen beruhen (Tolias et al. 2007; Ecker et al. 2007a, b), ist die Arbeit von mir selbst ausschließlich unter Einbeziehung der genannten Quellen und Hilfsmittel erstellt worden.

Tübingen, den 20. März 2008