Biol. Cybern. 88, 409–417 (2003)
DOI 10.1007/s00422-002-0387-1

Biological Cybernetics

Internal models for visual perception

Wolfram Erlhagen

Departamento de Matemática, Universidade do Minho – Campus de Azurém, 4800-058 Guimarães, Portugal

Received: 19 December 2001 / Accepted: 26 November 2002 / Published online: 3 April 2003

Abstract. Although the extrapolation of past perceptual history into the immediate and distant future is a fundamental phenomenon in everyday life, the underlying processing mechanisms are not well understood. A network model consisting of interacting excitatory and inhibitory cell populations coding for stimulus position is used to study the neuronal population response to a continuously moving stimulus. An adaptation mechanism is proposed that offers the possibility to control and modulate motion-induced extrapolation without changing the spatial interaction structure within the network. Using an occluder paradigm, functional advantages of an internally generated model of a moving stimulus are discussed. It is shown that the integration of such a model in processing leads to a faster and more reliable recognition of the input stream and allows for object permanence following occlusion. The modeling results are discussed in relation to recent experimental findings that show motion-induced extrapolation.

1 Introduction

Recent studies of localization errors observed with moving stimuli have caused an intense debate over the computational mechanisms underlying the encoding of position in the brain (for a recent review see Krekelberg and Lappe 2001). In the flash-lag illusion, observers are asked to judge the position of a continuously moving target relative to a stimulus briefly flashed in physical alignment with it. The moving target is perceived as leading the flash. It has been proposed that this localization error can be understood as the result of subthreshold and suprathreshold interactions between neuronal representations of stimulus frames along the motion trajectory (Kirschfeld and Kammer 1999; Sheth et al. 2000; Müsseler et al. 2002). In the explanation of Kirschfeld and Kammer, for instance, subthreshold activity from preceding frames facilitates a suprathreshold response to a direct stimulation. As a result, the delay to visual perception is reduced compared to the situation where the stimulus is flashed in isolation. Backward inhibitory forces from later frames, on the other hand, lead to a shortening of the response duration. Such explanations, which are essentially based on local cortical feedback, are consistent with the notion of an actively generated internal model of a moving target that is continuously compared with new afferent stimuli (Eagleman and Sejnowski 2000). Internal models have been proposed in the context of movement control as a mechanism to predict the consequences of our own action to maintain stability in the presence of feedback delays (for a review see Wolpert 1997). Similarly, compensating for neuronal delays between the retina and higher visual areas by extrapolating the position of a coherently moving object into the future was originally suggested by Nijhawan (1994) as an explanation for the flash-lag illusion. Although there is now experimental evidence that speaks against a complete compensation of transmission delays in the flash-lag paradigm (e.g., Whitney et al. 2000; but see Nijhawan 2001 and Khurana et al. 2000 for a different interpretation), there is evidence for extrapolatory perception under appropriate experimental conditions. When observers are asked to judge the vanishing point of a moving target, the perceived final position is typically displaced in movement direction (for a review see Hubbard 1995). More generally, extrapolating from perceptual history into the future is especially important whenever there is an interruption in the spatiotemporal continuity of the sensory input. In everyday life, objects move out of sight due to occlusion by other objects, but there is no doubt that we are able to infer their positions from past trajectory information.

In this article, I present a network model that aims to substantiate the notion of an internal model for visual perception and allows one to exemplify its functional consequences in the context of the processing of positional information. The key assumption is that the

---

Correspondence to: W. Erlhagen
(e-mail: wolfram.erlhagen@mct.uminho.pt)

internal model is triggered and may be influenced by external events but that its dynamic properties are essentially shaped by the excitatory and inhibitory feedback from within the network. In response to a moving stimulus, the network coding for position exhibits a spatiotemporal activity pattern consisting of a localized wave in visual field coordinates. We have recently used this dynamic model to explain the spatiotemporal characteristics of activity waves reconstructed in stimulus space from neuronal population activity in primary visual cortex (Erlhagen and Jancke 1999; Erlhagen and Jancke 2002, submitted). Here I broaden the scope of previous theoretical studies to experimental conditions in which the internal model does not directly interact with an afferent input. It is shown that a processing of positional information based on nonlinear cooperativity may lead to motion-induced extrapolation. Specifically, I focus on motion paradigms with occluding surfaces. Recent experimental findings both in electrophysiology and on the behavioral level indicate the existence of a neuronal representation of the moving target that persists also when the object is occluded (Assad and Maunsell 1995; Scholl and Pylyshyn 1999; Baker et al. 2001; Jellema and Perrett 2002). To cope with the problem of a temporary interruption of sensory information, a model architecture is proposed in which a stored internal model and the bottom-up input can be integrated.

The article is organized as follows: in Sect. 2 I briefly present the model architecture. Dynamic properties of the internal model, like subthreshold preshaping and extrapolation, are discussed in Sect. 3. Section 4 deals with the modeling work in occluder paradigms. The paper ends with a critical discussion of the modeling results in relation to recent experimental findings that show motion-induced extrapolation.

2 The model

The model network consists of an excitatory and inhibitory population of neurons that code for the position of a stimulus in visual space. Each neuron is parametrized by its visual receptive field center x. It is driven by external inputs representing afferents that carry information about the retinal location of the visual stimulus. The structure of the recurrent connections resembles a network architecture first studied by Wilson and Cowan (1973). Each excitatory neuron, x, integrates activity from neighboring neurons and projects via local connections to neurons of the inhibitory population that belong to the same functional column and via lateral connections to inhibitory neurons, x', with distinct visual receptive field centers ($x \neq x'$). Each inhibitory neuron is assumed to laterally integrate the incoming excitation but to project back to the excitatory population locally only (see Jancke et al. 1999 for details). For large neuronal populations the excitatory and inhibitory activity at time t in the network can be described by two continuous functions, $u(x,t)$ and $v(x,t)$, respectively (Amari 1977). The temporal evolution of these activities is governed by the following mean-field equations

$$\tau\dot{u}(x,t) = -u(x,t) + S(x,t) + h + g(u(x,t)) \times \left[\int w_u(x-x')f(u(x',t))dx' - v(x,t)\right]$$

$$\tau\dot{v}(x,t) = -v(x,t) + \int w_v(x-x')f(u(x',t))dx'$$

where $S(x,t)$ is the transient afferent input and $h < 0$ defines the resting level to which field activity relaxes without further stimulation. A Gaussian profile has been chosen for $S(x)$. Its amplitude, A_s, and space constant, σ_s, reflect the contrast and half-width of the localized stimulation. Throughout this paper all space variables are expressed as multiples of the basic unit $\sigma = \sigma_s$. The parameter τ defines the time constant of the dynamics. It is adjusted to reproduce the experimentally observed time scales. For the modeling of the population data in primary visual cortex, for instance, τ is assumed to be of the order of a few milliseconds. The response function f, which gives the expected portion of neurons at position x firing for a given level of excitation, is taken as a monotonically increasing function of typical sigmoid shape

$$f(u) = 1/(1 + \exp(-\beta(u - u_f)))$$

with threshold u_f and slope parameter β. The strength of the lateral interactions, $w_u(x,x')$ and $w_v(x,x')$, is assumed to decay as a function of the distance between sites x and x' in feature space. Gaussian decay functions with amplitude parameters A_u and A_v and spatial constants σ_u and σ_v are chosen for the excitatory and inhibitory population, respectively. In additon, as in many other field models it is assumed that the effective interaction strength in the network is inhibitory for sufficiently large separations between any two sites x and x' (e.g., Wilson and Cowan 1973; Ben-Yishai et al. 1997; Adini et al. 1997). Throughout this paper, all parameters describing the response function, f, and the interaction kernels, w_u and w_v, are kept constant. The values have been used in previous work to explain the spatiotemporal characteristics of population responses in primary visual cortex of cat (Jancke et al. 1999; Erlhagen and Jancke 2002, submitted). Finally, the interaction term in the excitatory field is multiplied by a nonlinear function $g(u)$ leading to a network of shunting type (for review see Grossberg 1988). $g(u)$ is assumed to be also of sigmoid shape with threshold u_g and slope β. Functionally, this state-dependent nonlinear signal serves to gate the lateral interactions by feedforward activation. For a sufficiently large threshold u_g the response properties of a neuron can only be influenced by interactions if the neuron receives direct bottom-up input. Physiological evidence supports such gating of lateral feedback. The spike discharge of most neurons in primary visual areas depend on contextual stimuli outside their receptive field only if the neuron is directly driven by another stimulus inside the receptive field (for a review see

Fitzpatrick 2000). Important to the present paper is the fact that the gating mechanism determines the extent to which the network dynamics may lead to an extrapolation of past trajectory information into the future (see below).

3 Spatiotemporal properties of the internal model for a moving stimulus

The presentation of a brief localized stimulus $S(x,t)$ of adequate intensity leads to a model response known as an active transient (Fig. 1A, Wilson and Cowan 1973). After stimulus offset the activity in the excitatory layer at stimulated sites continues to increase, reaches a maximum level, and then decays back to resting level due to the increasing inhibition in the network. Importantly, there is a threshold for the ignition of this active response. The external stimulation must be strong enough to trigger the self-stabilizing feedback loops. In the following, I refer to this threshold activity level as u_{TH} and to the corresponding stimulus intensity as A_{TH}. In response to a stimulus of adequate intensity that is continuously displaced along a horizontal line (apparent motion paradigm), the network develops a localized wave of activation (Fig. 1B). It follows the stimulus with a speed-dependent spatial lag (Erlhagen and Jancke 2002, submitted).

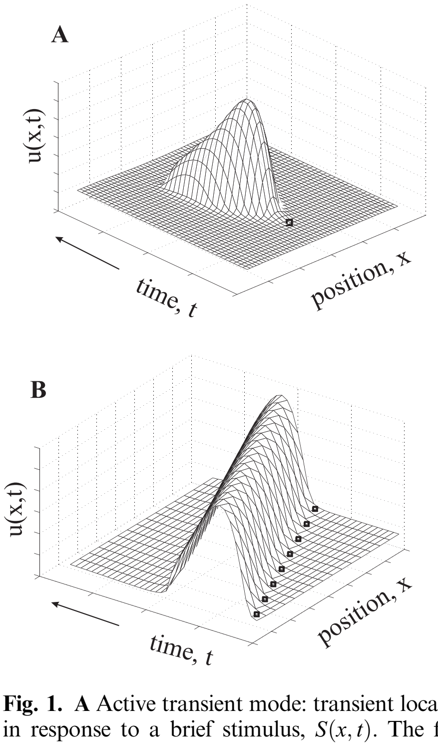

Fig. 1. **A** Active transient mode: transient localized activation pattern in response to a brief stimulus, $S(x,t)$. The field response $u(x,t)$ is centered over the stimulus position indicated by the *black square*. **B** Field response to a stimulus $S(x,t)$ that is continuously displaced by a distance Δx along a horizontal line. Each frame is presented for a time interval Δt leading to an apparent velocity $v = \Delta x/\Delta t$ of the stimulus train. *Black squares* indicate stimulus position; only a part in the middle of the trajectory is shown

3.1 Acceleration of processing through subthreshold preshaping

Due to the recurrent interactions, the response to a brief localized target stimulus is predicted to differ systematically in its spatiotemporal characteristics when it is preceded and succeeded by other stimuli compared to the situation when it is flashed in isolation (Fig. 2A). The temporal evolution of suprathreshold activity at the center position x_C of the stimulus in the stationary flash condition (solid line) and the moving condition is shown in Fig. 2B. Additionally, the response to the moving stimulus is compared for two stimulus trains with the same apparent velocity but achieved with different spatial displacements of the frames. In the spatially discrete case (dashed line) the displacement Δx is larger than the stimulus width $2\sigma_s$, whereas in the continuous case (dashed-dotted line) it is smaller. The latter case leads to a spatially overlapping stimulation and an earlier stimulation onset at position x_C compared to the discrete case and the flash condition. Importantly, in

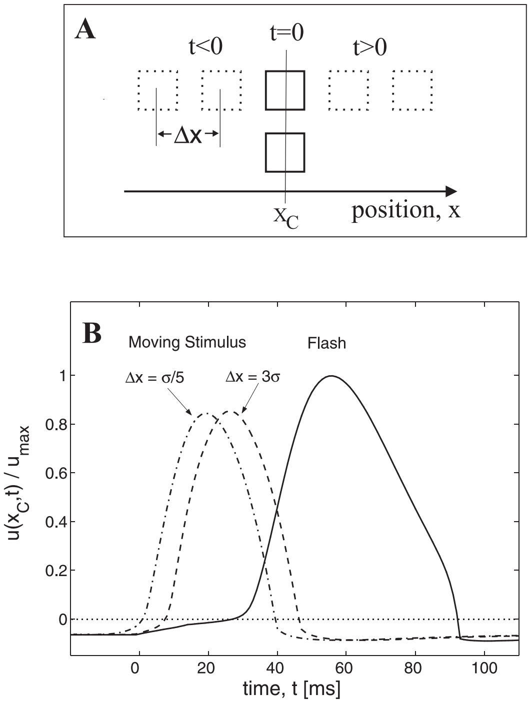

Fig. 2. **A** Sketch of the experimental setting in a flash-lag paradigm. A stationary stimulus centered over position x_C is flashed at time $t = 0$ together with a continuously displaced stimulus that is at that time in perfect vertical alignment with it. **B** Comparison of the field response at $x = x_C$ for the flash (*solid line*) and two motion displays in which the stimulus moves with the same apparent velocity $v = \Delta x/\Delta t$ but achieved with different displacements $\Delta x = \sigma/5$ (*dash-dotted line*) and $\Delta x = 3\sigma$ (*dashed line*) and accordingly adjusted frame durations Δt. The threshold activity u_{TH} is set to 0, that is, negative values of u indicate subthreshold activity. Stimulus parameters were: $A_s = 2.23$, $\sigma_s = \sigma = 10$ for the flash and the frames in the moving case, the flash duration $\Delta t = 15$ ms was chosen because it is equal to the frame duration for the the case $\Delta x = 3\sigma$. Model parameters were: $\tau = 15$ ms, $h = -3$, $\beta = 1.0$, $u_f = u_g = 0$, $A_u = 6.98$, $\sigma_u = 1.5\sigma$, $A_v = 7.98$, $\sigma_v = 2\sigma$

both moving cases the activity at x_C is below the threshold $u_{\mathrm{TH}} = 0$ when the flash is presented at time $t = 0$. However, when the direct stimulation sets in, the activity immediately bursts to a suprathreshold level, whereas the active response to the flash that lacks a subthreshold preactivation from preceding frames starts much later. Due to the inhibition from repesentations of lateral stimuli, the response at x_C in the moving condition does not reach the level of the flash representation and also decays earlier. As a result, the peak latency and the persistence of the suprathreshold response are both reduced in the order of 35 ms. When the flash response reaches its maximum, the representation of the moving stimulus will already be ahead of the flash position.

The loss of synchrony in the neuronal representations of simultaneous events in the outside world has been proposed as a fundamental principle to explain not only the flash-lag illusion (Kirschfeld and Kammer 1999; Whitney et al. 2000; for a detailed discussion with respect to the modeling result see Erlhagen and Jancke 2002, submitted) but also several other localization errors observed with moving retinal images (Schlag and Schlag-Rey 2002). On the other hand, the shortening of the response duration may functionally contribute to the "deblurring" of a moving object (Burr 1980).

3.2 Extrapolation

The simulation results in Fig. 2 illustrate that some of the processing delays are compensated for when a target stimulus is part of an apparent motion paradigm. However, the internal model still lags behind the actual position since the active response starts only upon target presentation. This does not mean, in general, that past trajectory information cannot be used by the visual system for a forward prediction. On the contrary, the dynamic properties of the proposed internal model of the moving stimulus suggest that the dynamic transformations may continue for a while also without further stimulation. The model simulation in Fig. 3A illustrates

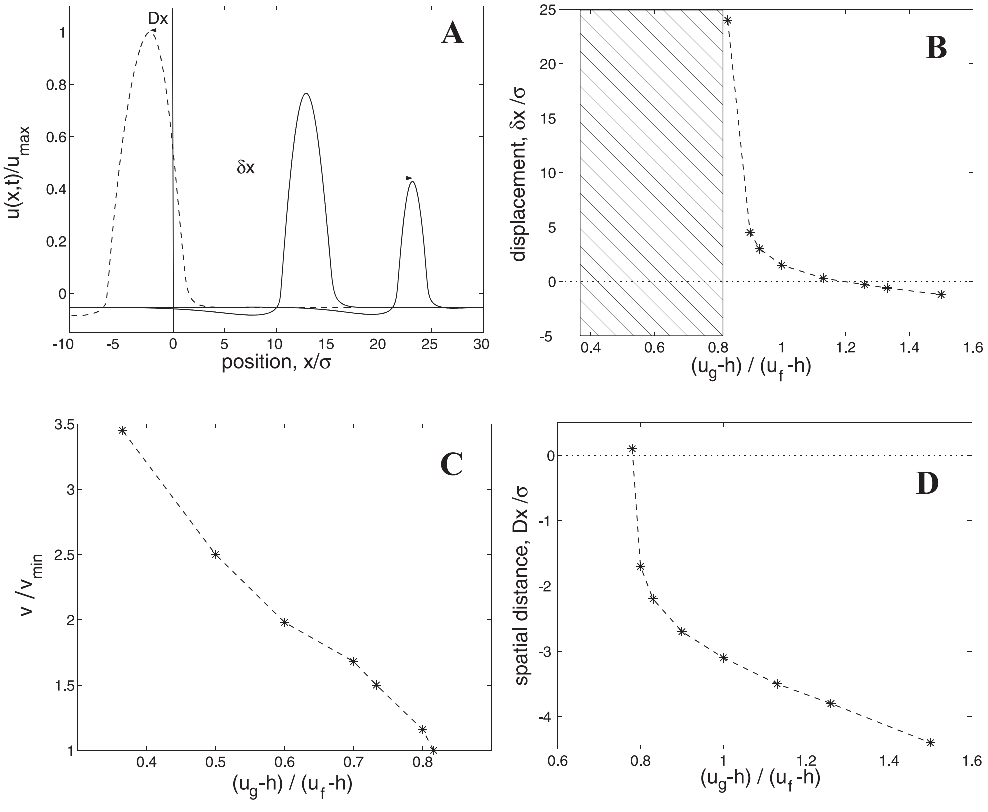

Fig. 3. **A** Three snapshots of the field activity $u(x, t)$ are shown in a condition in which the moving stimulus abruptly stops at position $x = 0$. The leftmost distribution (*dashed line*) represents the activity profile at the time of stimulation offset; the distance between its peak position and the vanishing point $x = 0$ is defined as the spatial lag Dx. The two other distributions represent the activity profile at two progressively later points in time. The forward displacement δx is defined as the distance between $x = 0$ and the peak position of the wave when it stops to travel and only decays back to resting level. Stimulus parameters were $\Delta x = 10$, $\Delta t = 5$ ms; $A_s = 4.0$. The shunting threshold was chosen as $u_g = -0.5$, and the rest of the model parameters were as in Fig. 2. **B** The spatial displacement is plotted against the dimensionless variable $(u_g - h)/(u_f - h)$, $(h = -3$ and $u_f = 0)$. The *hatched area* indicates the existence of self-sustained waves. **C** The velocity of the self-sustained wave relative to the minimum velocity, v_{min}, is plotted as a function of $(u_g - h)/(u_f - h)$. By adapting the threshold u_g it is possible to cover a velocity range of about $3.5 \times v_{min}$ without changes in the spatial interaction structure. Traveling waves with velocities beyond this range require a different scale of spatial cooperativity (for a discusion see Ermentrout 1998). Presumably the scale increases with increasing receptive field size from primary to higher visual areas. **D** The spatial lag Dx is plotted against $(u_g - h)/(u_f - h)$. A complete compensation of transmission delays $(Dx = 0)$ can be achieved for a sufficiently low shunting threshold u_g

that the neuronal representation, though progressively decreasing in amplitude, can reach positions far beyond the actual vanishing position. As mentioned earlier, a perceptual overshooting has been observed experimentally in tasks in which an observer must judge the vanishing point of a moving target without comparison stimulus. The forward displacement can exceed the stimulus dimension by several orders of magnitude (Hubbard 1995). In the network model, the value of the shunting threshold u_g modulates the extent to which an existing internal model of a moving stimulus may trigger a suprathreshold activation at positions devoid of afferent input. In Fig. 3B the forward displacement δx, measured as the distance from the actual vanishing point, is plotted against the dimensionless variable $(u_g - h)/(u_f - h)$. For values $u_g > u_f$ the displacement can become negative, indicating that the neuronal trajectory representation dies out before it reaches the last stimulated site. For smaller values $u_g < u_f$, however, displacements that are several orders of magnitude larger than the stimulus width can be observed. If the threshold of the gating mechanism is sufficiently low, the network model develops self-sustained waves that continue to travel in movement direction without further feedforward stimulation. The range of existence of this dynamic mode is indicated by the borders of the hatched area in Fig. 3b. Within this range the value of u_g affects the velocity of the traveling wave (Fig. 3C). As shown in Fig. 3D, the manipulation of a single model parameter, u_g, is sufficient to change from a representation that spatially lags behind the stimulus to a representation that is at or even slightly ahead of the stimulus position at the time of its presentation.

4 Internal models in the occluder paradigm

Probably one of the most prominent visual features in naturalistic scenes is a partial or complete occlusion of an object by other objects as they move in the visual field. For social interactions or in competitive situations it is advantageous not to disrupt the tracking of a moving object when the object disappears from view. There is recent experimental evidence both in neurophysiology and on the behavioral level that supports the notion of a neuronal trajectory representation that is not affected by brief interruptions in spatiotemporal continuity. Assad and Maunsell (1995) have reported cells in parietal cortex that increase their firing rate in blocks of "inferred motion" trials in which a stimulus disappeared without moving and then reappeared at a position consistent with the position where a continuously moving stimulus would be at that time. The activity was absent in control "blink" blocks where the stimulus reappeared at the vanishing point. More recently Perrett and colleagues (Baker et al. 2001; Jellema and Perrett 2002) have described a population of neurons, lying in an area of the temporal cortex associated with visual recognition, that shows increased activity when the experimenter or objects move behind an occluder and maximum activity following complete occlusion. The authors point out that the observed activity pattern is consistent with the idea that this neuronal population codes for the intention of the experimenter to go behind the occluder. Importantly, the presence of visible edges of occluding surfaces seems not to be necessary for the continuation of the internal representation. In a task in which observers had to track multiple moving items, Scholl and Pylyshyn (1999) have shown that the presence of either visible or virtual occluders does not disrupt the performance as long as the items disappear gradually from view.

The dynamic mode of a self-sustained traveling wave can, in principle, be used by the visual system to compensate for missing sensory information due to occluding surfaces. Figure 4 shows a model architecture in which an activity wave propagating with the velocity of the external stimulus train provides an additional top-down signal to the excitatory population of the position field.

Such an architecture is in line with a growing body of anatomical and psychophysical evidence suggesting that top-down effects play a crucial role in the processing of input stimuli (for a recent review see Lamme and

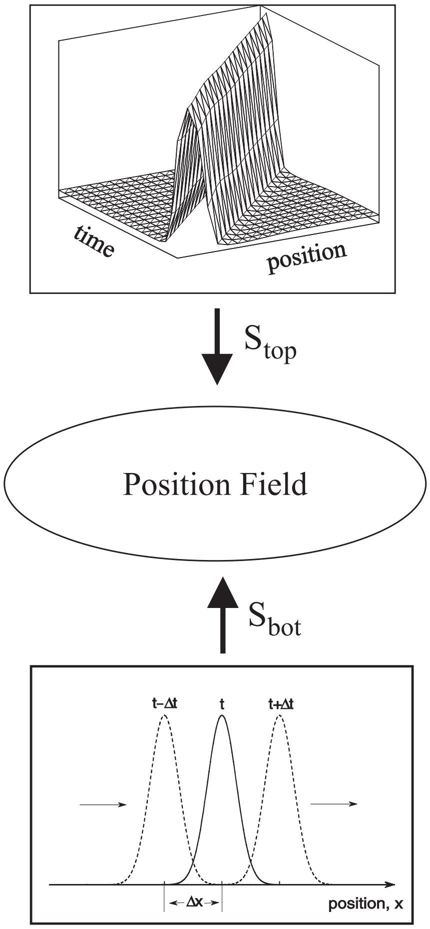

Fig. 4. Sketch of the model architecture. The position field represented by the inhibitory and excitatory neuronal populations gets bottom-up as well as top-down signals. The top-down contribution S_{top} from higher visual areas consists of a traveling wave representing a self-stabilized internal model for the position of a moving object. The bottom-up input, S_{bot}, represents an apparent motion display in which a stimulus is successively displaced along a horizontal line

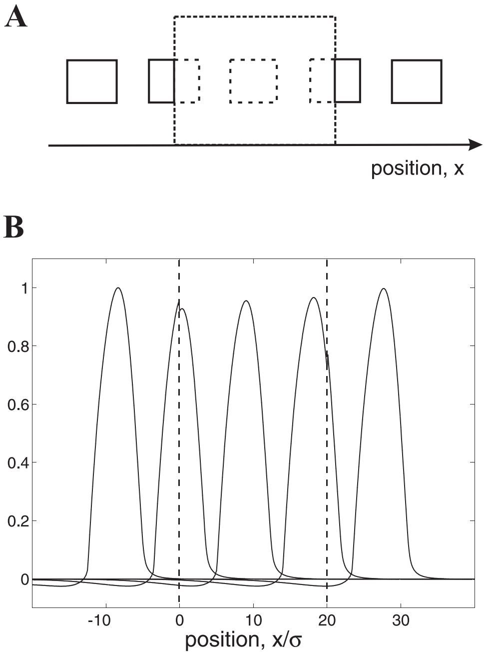

Fig. 5. **A** Sketch of a virtual occluder paradigm. A stimulus is continuously displaced along a horizontal line from left to right (not all frames are shown to avoid crowding). At the left border of the virtual occluder (*large rectangle*) the stimulus gradually disappears and reappears at the opposite side. **B** Five snapshots of the field response $u(x,t)$ triggered by the total input $S(x,t) = S_{bot}(x,t) + S_{top}(x,t)$ to the excitatory population are shown. The borders of the virtual occluder of width 20σ are indicated by the *dashed lines*. All activity profiles are separated by a constant time interval of 45 ms. The top-down input consisted of a self-sustained traveling wave that was scaled to 95% of the threshold amplitude, A_{TH}, necessary to trigger an active response. The choice of the value $u_g = -0.65$ ensured that the wave traveled with the apparent velocity of the external stimulus train ($\Delta x = 10$, $\Delta t = 5$). Note that the profile at about $x = 10\sigma$ represents a part of the trajectory without any bottom-up signal. Parameters of the bottom-up stimulus were $A_S = 4.0$, $\sigma_s = 10$; model parameters were as in Fig. 2

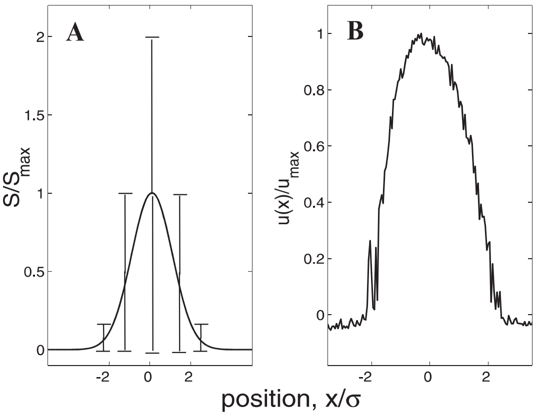

Fig. 6. Disambiguation of weak and noisy bottom-up information by a top-down priming process. **A** A single frame of the noisy bottom-up input stream is shown along with error bars ($A_s = 0.96$, $\sigma_s = 10$, $\Delta x = 10$, $\Delta t = 5$). This stimulus train does not trigger an active field response, that is, $A_s < A_{\mathrm{TH}}$. **B** The integration of a corresponding top-down input leads nevertheless to a suprathreshold field reponse. The plotted snapshot of the traveling wave indicates a much improved signal-to-noise ratio. The top-down input and the model parameters were as in Fig. 5

Roelfsema 2000). It is further assumed that this additional top-down input from a higher visual area does not directly drive the cells but only leads to a subthreshold preactivation of the excitatory population. This assumption coincides with results of physiological studies that report a rather subtle modulator effect of feedback connections on receptive field properties (Lamme and Roelfsema 2000). The assumed "priming-like" effect is, however, sufficient to maintain the population activity at a suprathreshold level following occlusion. Figure 5 shows five snapshots of a model simulation in an occluder paradigm. Those portions of the continuously displaced stimulus that happened to intersect with a virtual rectangle of width 20σ are set to zero (compare the sketch in Fig. 5A). The sequence of profiles at successively later points in time illustrates that the amplitude of the representation dips slightly when the border of the virtual rectangle (dashed line) is crossed but that the suprathreshold wave continues tracking the position of the occluded stimulus (Fig. 5B).

As shown in Fig. 6, a hypothesis about the actual stimulus train expressed by the higher area's activity also leads to an efficient disambiguation of input data corrupted by noise. In Fig. 6A, a single frame of a bottom-up stimulus train used in the numerical studies is plotted along with error bars. They indicate a noise level equal to the mean value. This bottom-up input S_{bot} alone does not trigger an active response, that is, the input strength does not exceed the threshold A_{TH}. However, if S_{bot} is primed by providing corresponding top-down signals, S_{top}, a coherently traveling wave evolves (compare the snapshot in Fig. 6B). Due to the lateral interactions in the network, this suprathreshold population response to the combined input, $S(x,t) = S_{bot}(x,t) + S_{top}(x,t)$, shows a much improved signal-to-noise ratio (Douglas et al. 1995; Salinas and Abbott 1996).

It is important to note that the two subthreshold input streams resonate only when they are sufficiently similar in their spatiotemporal characteristics. The degree of match between the corresponding patterns necessary to trigger a suprathreshold wave has been systematically analyzed in Fig. 7. The "allowed" maximum difference in velocity, $\Delta v = |v_{top} - v_{bot}|$, is plotted against the relative strength A_{top}/A_{TH} of the subthreshold top-down signal. If A_{top} is close to the threshold A_{TH}, the velocity of the stimulus train can exceed that of the internal model by more than 150% and vice versa. A sufficiently weak input, S_{top}, on the other hand, requires a nearly perfect match of the two velocities to trigger the suprathreshold wave. A similar relation holds also for an assumed temporal offset between the internal model and the bottom-up signal (not shown). A decrease in the strength of the top-down contribution can thus be used to successively increase the degree to which the spatiotemporal characteristics of the two input patterns must coincide. Given a certain bottom-up pattern, a more demanding degree of match possibly requires a (modest) change in the velocity of the internal model. Such a refinement can be achieved without changes in the spatial

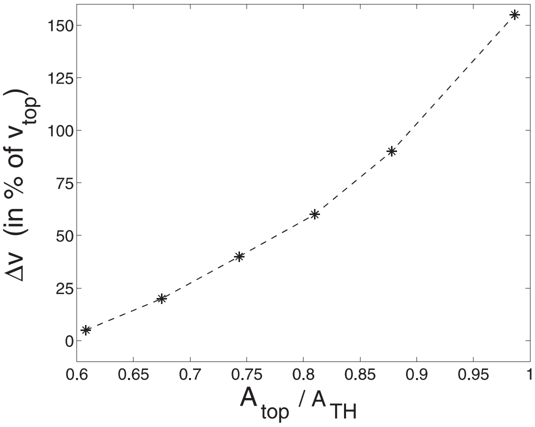

Fig. 7. The degree of match between the two subthreshold inputs to the position field in Fig. 6 necessary to trigger a suprathreshold reponse has been analyzed. The maximum difference $\Delta v = |v_{bot} - v_{top}|$ of the two velocities is plotted as a function of the relative strength A_{top}/A_{TH} of the top-down signal. The velocity of the top-down signal was kept fixed and the velocity of the bottom-up input was changed by manipulating the spatial distance Δx between successive frames of the stimulus train. In general, the relative stronger input, S_{bot} or S_{top}, determines the velocity of the evolving wave

interaction structure by adapting the shunting threshold u_g (compare Fig. 3C).

The described resonance phenomenon can be exploited to learn for future use the link between bottom-up input streams and their corresponding internal models. It is assumed that during learning and development successful combinations of S_{bot} and S_{top}, which lead to a suprathreshold response in the position field, are reinforced. A simple correlational learning rule would be sufficient to establish a direct (or indirect) mapping between the corresponding neuronal populations. As a consequence, bottom-up information resulting from similiar input patterns flows up along the ascending pathway and selects automatically the corresponding internal model represented in the higher visual area.

4.1 Anticipation

As shown in the model simulations (Fig. 3d), the processing mechanism of nonlinear cooperativity may compensate for visual delays. The neuronal population representation of the moving stimulus can be slightly ahead of the actual stimulus position. However, recordings from neurons in several brain areas suggest that internal brain states like intention, expectation, or prior experience may result in neuronal responses that anticipate sensory changes by far more than a few milliseconds (e.g., Duhamel et al. 1992; Umiltà et al. 2001). Assad and Maunsell (1995) report a significant increase in the neuronal population signal in parietal cortex up to 120 ms before the disappearence of the target in trials that allow the inference that the invisible target is moving. On the assumption that the internal model can be triggerd before the presentation of the actual stimulus train, the model architecture with bottom-up as well as top-down input to the position field accounts for a larger anticipatory effect. Figure 8 illustrates that a lagging population response to the bottom-up input starts to lead the actual stimulus position if the additional top-down signal precedes the onset of the bottom-up stimulation by 15 ms. The spatial lead becomes continuously larger with increasing temporal offset. The maximum offset, which still results in anticipation, critically depends on the temporal dynamics and the spatial extent of the subthreshold priming by the top-down signal (see Discussion).

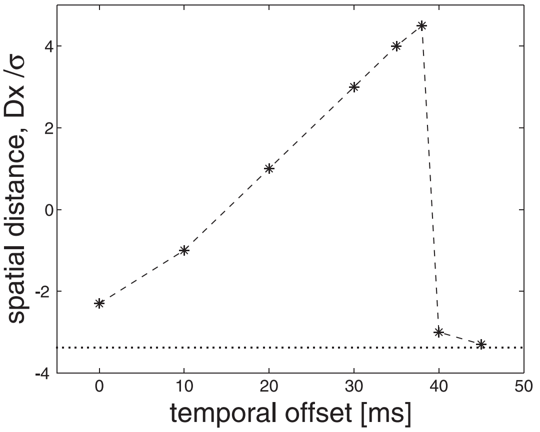

Fig. 8. An assumed temporal lead of the top-down input affects the spatial distance, *Dx*, between the actual stimulus position and the peak of the population representation. For the present parameter setting, a spatial lag turns into a spatial lead if the top-down signal preceds the bottom-up input stream by more than 15 ms. Beyond a temporal offset of 40 ms the two signals do not interact anymore. As a consequence, the population representation lags behind the actual stimulus position, as would be predicted without an additional priming process (*dotted line*). Note that the present choice of $u_g = 0.8$ does not lead to a complete compensation of transmission delays. The top-down input, the bottom-up input, and the rest of the model parameters were as in Fig. 5

5 Discussion

In this article, it is suggested that the perception of position of a moving object is the result of an interplay between the actual sensory input and an internal modeling process. Nonlinear interaction processes triggered by the input stream lead to a wavy activity pattern in the network model. This wave is interpreted as a neuronal representation for the prediction of stimulus position. The network architecture, with its pattern of connectivity, resembles recurrent networks that have been proposed for the processing of visual information several times in the past (e.g., Wilson and Cowan 1973; Ben-Yishai et al. 1997; Adini et al. 1997). In contrast to these theoretical studies, the aim of the present work is to extend model predictions to experimental conditions in which no direct link between sensory and perceptual events exists. A second major difference is the emphasis on subthreshold contributions for the processing of visual information.

There are two main findings of this work. First, in an appropriate parameter regime the wave exhibits properties like extrapolation and anticipation that are usually attributed to the notion of an internal model for the prediction of future states of the environment. Second, the priming of the actual stimulus by an internal model stored in a higher visual area leads to a faster and more reliable interpretation of the bottom-up input and allows for object permanence in occluder paradigms.

The key question addressed in a host of studies with moving stimuli following Nijhawan's report in 1994 is whether the visual system uses spatial extrapolation based on prior trajectory information to compensate for processing delays (Krekelberg and Lappe 2001). The modeling results show that nonlinear cooperativity among cortical neurons may lead to a trajectory representation that does not necessarily lag behind the actual stimulus position. In addition, a recent report about an anticipatory wave of spiking activity in a population of retinal ganglion cells (Berry et al. 1999) indicates that mechanisms for motion-induced extrapolation may be found at different stages of the visual pathway. However, whether a complete compensation of transmission delays occurs under the conditions of the flash-lag paradigm is still a matter of debate (Nijhawan 2001; Khurana et al. 2000). In the model, a "correction" to match the object's actual position requires a strong influence of recurrent interactions on the response of the neuronal population. Cooperativity in the network tends, however, to stabilize the traveling wave also after cessation of the moving stimulus. A direct experimental test should thus include a "stopping condition" in which the accompanying flash is presented with a sufficient delay upon the offset of the moving target. On the assumption that the compensation is complete, the model predicts a judgement bias in movement direction. Assuming, on the other hand, that the population representation is spatially lagging and does not appear to overshoot the vanishing point (Whitney et al. 2000; Müsseler et al. 2002) requires an explanation of the observed localization error under absolute judgement condition. Without accompanying flash a clear localization error in the direction of movement is observed (Müsseler et al. 2002). It is suggested that the prior knowledge about the task setting already defines how predictive the self-stabilized internal model is. It should be less predictive under relative judgement condition since then the task is not to optimize the processing of subsequent stimulus frames but to allow for comparison with a fixed but unpredictable position in space. In the model, the pattern of displacement at the end of the movement trajectory is determined by the threshold of the gating mechanism (see Fig. 3B). A dependence of the shunting term on the summed activity, $\sum_i u_i$, of a neuronal population representing the expected occurrence of a task-relevant flash, that is, $g(u, \sum_i u_i) = g(u + \sum_i u_i)$, may shift the threshold adequately. There is substantial evidence from physiological studies on contextual effects and the effect of attentional demands that supports the notion of such a task-dependent adaptive gating of recurrent processing (Lamme and Roelfsema 2000).

The idea that a stored internal model is used by the visual system to predict, via feedback connections, bottom-up information has been previously addressed by several authors (Grossberg 1980; Mumford 1992; Rao and Ballard 1997). In these studies, the assumed bidirectional processing scheme requires a sophisticated pattern-matching mechanism. The patterns carried by the top-down and the bottom-up pathway must be compared and the remaining difference projected back to the higher area to modify the internal model. In the work presented here, the threshold behavior of neurons is exploited to establish a link between the two signals. The priming of the position field by the top-down input resonates with the bottom-up input only if the velocities of the two input streams are sufficiently similiar. This resonance phenomenon allows for the simultaneous exploration of multiple internal models covering distinct velocity ranges without the assumption of an additional competition process. Such a parallel search in multiple top-down, bottom-up loops is a basic characteristics of Ullman's model architecture for visual processing (Ullman 1994).

The present implementation obviously simplifies various aspects of the integration of the top-down and the bottom-up input. Anatomical data suggest that the two types of signals are mapped on separate but interacting neuronal populations (Lamme and Roelfsema 2000). Second, neuronal population activity as a level of description does not address the cellular mechanisms underlying this integration process. The results of a recent modeling study by Siegel et al. (2000) suggest that somatodendritic interactions could play an important role. A further simplification concerns aspects of the spatiotemporal characteristics of the top-down priming that directly affect the magnitude of model properties such as anticipation. A scaled version of a self-sustained suprathreshold wave has been used in the simulations as a subthreshold top-down input (compare Fig. 5). Up to now, no experimental study has directly addressed the time course and spatial extent of the subthreshold response to a motion display. However, intracellular recordings and optical imaging in primary visual cortex suggest that the spread is much larger than would be predicted by the spike activity of a neuronal population (Bringuier et al. 1999; Toth et al. 1996). Preliminary model simulations show that an assumed larger spread of the preactivation results in a larger anticipation of the actual stimulus position compared to the model predictions in Fig. 8.

The modeling results lead to several experimental predictions. Most importantly, it is shown that the subthreshold top-down signal may result in a strong motion-induced extrapolation. This happens to occur even though the recurrent interactions at the lower area do not support this property. In the simulation shown in Fig. 6, model parameters have been used that were adapted to explain population responses to motion displays in primary visual cortex of anesthetized cat (Erlhagen and Jancke 2002, submitted). Without priming, the population representation is predicted to only slightly overshoot the point of disappearence, that is, the border of the occluder. However, the additional priming input is sufficient to maintain the traveling wave following occlusion. This model prediction could be directly tested in paradigms in which the animal is forced to actively infer the position of a temporary hidden object. Tanscranial magnetic stimulation could then be used to silence the assumed top-down signal form higher visual areas for a short period of time.

It would be interesting to reconstruct the motion trajectory in an occluder paradigm from population activity in temporal cortex with the same estimation techniques we have used for the population data in primary visual areas (Jancke et al. 1999; Erlhagen et al. 1999). Baker and coworkers (2001) report a subpopulation of position-specific "occluder cells" that increase their activity under gradual occlusion condition and not when the target abruptly disappears. A trajectory reconstruction based on a whole population of position-specific occluder neurons is predicted to provide insight into how far the population activity propagates under different conditions of disappearence. The modeling results suggest a continuum between a perfect extrapolation and a relatively fast decay of the population activity to resting level.

Acknowledgements. The author would like to thank D. Jancke for useful discussions and two anonymous reviewers for helpful comments and suggestions on a previous version of this paper. This research was supported by a European grant (IST-2000-29689) and by the Portuguese Science Foundation (POSI/SRI/38051/2001).

References

Adini Y, Sagi D, Tsodyks M (1997) Excitatory-inhibitory network in the visual cortex: psychophysical evidence. Proc Natl Acad Sci USA 94: 10426–10431

Amari S (1977) Dynamics of pattern formation in lateral-inhibitory type neural fields. Biol Cybern 27: 77–87

Assad JA, Maunsell JHR (1995) Neuronal correlates of inferred motion in primate posterior parietal cortex. Nature 373: 518–521

Baker CI, Keysers C, Jellema T, Wicker B, Perrett DI (2001) Neuronal representation of disappearing and hidden objects in temporal cortex of macaque. Exp Brain Res 140: 375–381

Ben-Yishai R, Hansel D, Sompolinsky H (1997) Traveling waves and the processing of weakly tuned inputs in a cortical network module. J Comput Neurosci 4: 57–77

Berry MJ II, Brivanlou IH, Jordan TH, Meister M (1999) Anticipation of moving stimuli by the retina. Nature 398: 334–338

Bringuier V, Chavane F, Glaeser L, Fregnac Y (1999) Horizontal propagation of visual activity in the synaptic integration field of area 17 neurons. Science 283: 695–699

Burr DC (1980) Motion smear. Nature 284: 164–165

Douglas RJ, Koch C, Mahowald M, Martin KAC, Suarez HH (1995) Recurrent excitation in neocortical circuits. Science 269: 981–985

Duhamel J, Colby CL, Goldberg ME (1992) The updating of the representation of visual space in parietal cortex by intended eye movements. Science 255: 90–92

Eagleman DM, Sejnowski TJ (2000) Reply to Krekelberg et al. Science 289: 1107

Erlhagen W, Bastian A, Jancke D, Riehle A, Schöner G (1999) The distribution of neuronal population activation as a tool to study interaction and integration in cortical representations. J Neurosci Meth 94: 53–66

Erlhagen W, Jancke D (1999) Motion waves in primary visual cortex as a neural correlate for the perception of moving objects. Abstr Soc Neurosci 25: 679

Erlhagen W, Jancke D (2002) A theory for the processing of position in visual space

Ermentrout B (1998) The analysis of synaptically generated traveling waves. J Comput Neurosci 5: 191–208

Fitzpatrick D (2000) Seeing beyond the receptive field in primary visual cortex. Curr Opin Neurobiol 10: 438–443

Grossberg S (1980) How does a brain build a cognitive code? Psychol Rev 87: 1–51

Grossberg S (1988) Nonlinear neural networks: Principles, mechanisms, and architectures. Neural Netw 1: 17–61

Hubbard TL (1995) Environmental invariants in the representation of motion: Implied dynamics and representational momentum, gravity, friction, and centripetal forces. Psychon Bull Rev 2: 322–338

Jancke D, Erlhagen W, Dinse HR, Akhavan AC, Giese M, Steinhage A, Schöner G (1999) Parametric population representation of retinal location: neuronal interaction dynamics in cat primary visual cortex. J Neurosci 19: 9016–9028

Jellema T, Perrett DI (2002) Coding of visible and hidden actions. In: Prinz W, Hommel B (eds) Attention and performance, XIX. Oxford University Press

Khurana B, Watanabe K, Nijhawan R (2000) The role of attention in motion extrapolation: are moving objects 'corrected' or flashed objects attentionally delayed? Perception 29: 675–692

Kirschfeld K, Kammer T (1999) The Fröhlich effect: a consequence of the interaction of visual focal attention and metacontrast. Vision Res 39: 3702–3709

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

Lamme VAF, Roelfsema R (2000) The distinct modes of vision offered by feedforward and recurrent processing. Trends Neurosci 23: 571–579

Mumford D (1992) On the computational architecture of the neocortex. II. The role of cortico-cortical loops. Biol Cybern 66: 241–251

Müsseler J, Storck S, Kerzel D (2002) Comparing mislocalizations with moving stimuli: the Fröhlich effect, the flash-lag, and representational momentum. Vis Cognition 9(1/2): 120–138

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

Nijhawan R (2001) The flash-lag phenomenon: object motion and eye movements. Perception 30: 263–282

Rao RPN, Ballard DH (1997) Dynamic model of visual recognition predicts neural response properties in the visual cortex. Neural Comput 9: 721–763

Salinas E, Abbott LF (1996) A model of multiplicative neural responses in parietal cortex. Proc Natl Acad Sci USA 93: 11956–11961

Schlag J, Schlag-Rey M (2002) Through the eye slowly: delays and localization errors in the visual system. Nat Rev 3: 191–200

Scholl BJ, Pylyshyn ZW (1999) Tracking multiple items through oclusion: clues to visual objecthood. Cognit Psychol 38: 259–290

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

Siegel M, Köding KP, König P (2000) Inegrating top-down and bottom-up sensory processing by somato-dendritic interactions. J Comput Neurosci 8: 161–173

Toth LJ, Chenchal R, Dae-Shik K, Somers D, Sur M (1996) Subthreshold facilitation and suppression in primary visual cortex revealed by intrinsic signal imaging. Proc Natl Acad Sci USA 93: 9869–9874

Ullman S (1994) Sequence seeking and counterstreams: a model for the bi-directional information flow in the cortex. In: Koch C, Davis J (eds) Large-scale theories of the cortex. MIT Press, Cambridge, MA, pp 257–270

Umiltà MA, Kohler E, Gallese V, Fogassi L, Fadiga L, Keysers C, Rizzolatti G (2001) I know what you are doing: a neurophysiological study. Neuron 31: 155–165

Whitney D, Murakami I, 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 Res 40: 137–149

Wilson HR, Cowan JD (1973) A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. Kybernetik 13: 55–80

Wolpert DM (1997) Computational approaches to motor control. Trends Cogn Sci 1: 209–216