A neural model of feature attention in motion perception

Pierre Bayerl* , Heiko Neumann1

University of Ulm, Department of Neural Information Processing, Oberer Eselsberg, D-89069 Ulm, Germany

Received 24 November 2005; accepted 18 April 2006

Abstract

We utilize a model of motion perception to link a physiological study of feature attention in cortical motion processing to a psychophysical experiment of motion perception. We explain effects of feature attention by modulatory excitation of neural activity patterns in a framework of biased competition. Our model allows us to qualitatively replicate physiological data concerning attentional modulation and to generate model behavior in a decision experiment that is consistent with psychophysical observations. Furthermore, our investigation makes predictions for future psychophysical experiments.

© 2006 Elsevier Ireland Ltd. All rights reserved.

Keywords: Feature attention; Biased competition; Motion perception; Neural modeling

1. Introduction

Cortical motion processing is confronted with the problem of reliably estimating motion cues and to integrate them into consistent object related interpretations. Processes of attention help to increase the separation of localized features in order to group them together forming coherent object motions. In this work we build upon and further extend a neural model of cortical motion perception to link the outcome of experimental studies resulting from different investigations. Our studies combine experimental evidence concerning feature attention from electrophysiological and psychophysical observations. The proposed model makes further testable predictions that can be verified in future experiments.

1.1. Feature attention

Feature-based attention in early vision describes the deployment of attentional load to a specific feature irrespective of its spatial location, such as, e.g., motion direction (Martinez-Trujillo and Treue, 2004; Treue and Martinez-Trujillo, 1999) or orientation (Reynolds and Chelazzi, 2004). This type of attention selection is distinguished from spatial, or location-based, attention where information is expected at some location irrelevant of the feature and from object-based attention that is assumed to operate on chunks of already grouped features that form individual objects (Blaser et al., 2000). In this work we focus on feature attention.2

In the following we outline two experimental studies, which investigate the effect of feature attention in motion perception.

* Corresponding author. Tel.: +49 731 50 24157; fax: +49 731 50 24156.
E-mail addresses: pierre.bayerl@uni-ulm.de (P. Bayerl), heiko.neumann@uni-ulm.de (H. Neumann).
1 Tel.: +49 731 50 24158; fax: +49 731 50 24156.

2 Furthermore, our modeling framework also allows spatial attention to be included (not further investigated in this paper).
1.2. Relevant experimental studies

It has recently been demonstrated by Martinez-Trujillo and Treue (2004) that in area MT of the macaque cell responses are modulated by feature-based mechanisms of attention. The authors evaluated cell responses to different directions of motion with different configurations of feature attention: (1) no attention to any specific motion; (2) attention towards the velocity that is preferred by the investigated cell; (3) attention towards the presented velocity. These conditions lead to the following key observations:

- For the experimental condition with feature attention towards the preferred direction of the observed cell an increased response was observed compared to the condition without attention independently of the presented motion pattern.
- When attention was directed to the presented motion two cases have to be distinguished: (1) an increased response of the investigated cell was observed compared to the condition without attention when the presented motion pattern was identical or similar to the preferred motion of the observed cell; (2) a decreased response was observed, on the other hand, when the direction of motion was very dissimilar to the preferred direction of motion of the observed cell.

A quantitative evaluation of their data showed that investigated cells are modulated by feature attention. Consistently, in the presence of modulatory feature attention towards the preferred direction of motion the shape of the cells' tuning curves did not significantly change.

The psychophysical study of Felisberti and Zanker (2005) demonstrated that attention significantly enhances the ability of humans to detect motion in transparent motion patterns. For a given number of transparent layers of motion, the authors compared the threshold at which a specific motion can reliably be detected for different conditions of feature attention. The experiments utilized an alternative forced choice task in which the observers had to decide about the presence or absence of a specific direction of motion. In one condition, the subjects were not informed about the motion direction prior to stimulus presentation, whereas in the other condition they were. This additional information guides the feature attention. The obtained results show an increased threshold (indicating more distinguishable transparent layers) when attention was directed to the direction of interest.

1.3. Modeling approach and outline of this work

In the next section we outline the main features of our model. First, we present an introductory example processing a motion sequence demonstrating the effect of feature attention. Then, systematic computational simulations are presented for a set of stimuli with moving random dots containing motion in different directions. These results are related to the experimental data presented in Martinez-Trujillo and Treue (2004) and Felisberti and Zanker (2005). Our simulations suggest an interpretation of the observed data in the framework of biased competition (Reynolds and Chelazzi, 2004), which is generated here by mechanisms of modulatory enhancement of activation (Eckhorn et al., 1990; Neumann and Sepp, 1999) and lateral competitive processing (Simoncelli and Heeger, 1998). The employed model is based on our previous work (Bayerl and Neumann, 2004) with the extension that global feature attention is included in the model dynamics to provide a top-down modulation signal.

2. Model

We model motion sensitive cells in two areas of the dorsal visual stream, namely areas V1 and MT. In addition to these model areas we define a fixed task-related attentional signal that is fed into the integration scheme defined by bidirectional connections between model areas V1 and MT. The dynamics of each model area is defined by feedforward integration, lateral inhibition, and top-down feedback modulation (the equations and the parameters used for simulations are given in the Appendix A). The input to the model, and thus the input to model area V1, is generated by a correlation detector operating on band-pass filtered frames of the input sequence (see Appendix A).

2.1. Feedback and attentional modulation

Treue and Martinez-Trujillo suggested that attention modulates the activities in early motion areas (Martinez-Trujillo and Treue, 2004; see also Treue and Martinez-Trujillo, 1999). We thus model feature attention by utilizing a modulatory feedback signal to motion sensitive cells (see Fig. 1 for a sketch

---

3 In our model all motion sensitive cells are tuned to speed as well as to motion direction (Bayerl and Neumann, 2004). Thus, each motion sensitive cell in our model is tuned to one specific “preferred velocity” or “preferred motion”. Depending on the context or the experimental condition we will also use the term “preferred direction” if the preferred speed is not of interest.
Fig. 1. Model sketch illustrating feedback connections between the two model areas (V1 and MT) of the presented model. Feedback enhances the input to the model areas in an excitatory manner by modulating input activities with a feedback signal larger or equal to one. Thus, the input activities cannot directly be suppressed by the feedback signal, nor can feedback induce activity without any feedforward input present. (1) Feature attention modulates activities representing a specific motion feature (velocity) in model MT irrelevant of the spatial location. (2) Local expectations are computed in model MT by spatial pooling and are utilized to enhance input activities to model V1 matching these local expectations.

describing feedback modulation). This signal globally enhances those activities in model area MT, which represent the attended global motion feature: input(to MT)·(1 + attention). Importantly, when no attention is given (zero valued attention signal) the input is left unchanged. This allows feature signals to survive even if they were momentarily not enhanced by such feedback signal. This is different from mechanisms in which the input gates the feedback and where, as a consequence, an unattended feature response is extinguished and cannot be restored for later processing again. The same feedback mechanism is utilized to model early feedback between model areas MT and V1 to locally enhance expected local motion patterns while leaving other activities unchanged: input(to V1)·(1 + expectation(from MT)), see (Fig. 1 and Bayerl and Neumann, 2004). Such modulations in early parts of the dorsal stream have been reported on the basis of experimental investigations by, e.g., Friston and Büchel (2000), Hupé et al. (2001). In the presented model simulations (see results) global expectations concerning feature attention are task-related and given by some external signal. Alternatively, it would be possible to add additional stages to process the outcome of model MT and/or other areas to compute the expectation by combining bottom up input. Local top-down expectations are generated in model area MT. Here activity patterns from model V1 are integrated in a small spatial neighborhood which, in turn, define the average velocity patterns that are used as a prediction to evaluate incoming stimulus data over time (see Appendix A, compare model parameterization in Table 1).

2.2. Lateral inhibition

Within each area the modulated input activities are squared (employing a nonlinear transfer-function) which is supposed to enhance differences between activities. To keep the modulated input activities in certain bounds we apply a local lateral shunting inhibition scheme to preserve the total energy of activity patterns at each location. This normalization leads to large activities only if few motion cues are presented at isolated locations (e.g., opaque motion) and to smaller activities in the presence of multiple concurring motion cues (e.g., in the presence of transparent motion).

2.3. Biased competition

Together, feedback modulation and feature competition (normalization) lead to a biased competition. Attention and early feedback iteratively influence the neural activity balance in both model areas by enhancing expected patterns at the cost of decreasing locally the activity of unexpected motion patterns through lateral normalization.

2.4. Example

Fig. 2 shows the relaxed output of model area MT processing a motion sequence showing three objects defined by random dots moving in front of a moving background. The task of the observer in this case could be to extract the object moving to the left. This output of the model simulation illustrates the effect and the use of feature attention. The maximum activity of cells representing objects with the attended motion feature (here: the object

<table>
<thead>
<tr>
<th>Table 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model parameterization</td>
</tr>
<tr>
<td>netIN</td>
</tr>
<tr>
<td>Model V1</td>
</tr>
<tr>
<td>Model MT</td>
</tr>
</tbody>
</table>

Compared to the presented model, in the original model (Bayerl and Neumann, 2004) netFB and A were set to zero for model MT.
Fig. 2. Example processing an example sequence to illustrate the effect of feature attention. Top row: model sketch with and without feature attention (left and right). One frame from the input sequence (center) showing random dots. Circular objects are defined by visual motion in horizontal directions on a vertical moving background. For illustration purposes, objects are indicated by dotted circles. The true motion of the objects presented in the stimuli is illustrated by black arrows. Bottom row: relaxed model results after eight iterations of feedback processing with and without feature attention (left and right). The maximum activity of model MT cells over all velocities is shown at each location. Brightness indicates activity of the cell encoding the most active velocity. Motion vectors corresponding to the tuning of the cell with the maximum activity are shown at selected locations. Left: without attention motion is detected segmenting the scene in regions corresponding to differently moving objects. Right: with attention to medium leftward motion (−3 pixel/frame) the image region moving with the attended velocity is highlighted. Thus, feature attention may help finding objects with certain features in the presence of many other objects with other features.

in central vertical position) is highlighted, which may help finding this specific object in the presence of other moving objects or clutter.

3. Simulations and results

3.1. Stimuli

The stimuli (80 × 80 pixel) used in this section consist of image sequences showing white random dots on black background moving in different directions. The number of directions presented to the model (a) is 1 for the experiment related to Martinez-Trujillo and Treue (2004) and (b) varies from 1 to 7 for the experiment related to Felisberti and Zanker (2005).

In the following simulations we investigate the response of one individual cell in the output stage of model MT with its receptive field located at the center of the visual field and a selectivity tuned to one specific velocity depending on the experimental configuration.

3.2. Influence of attention on the response of motion cells to opaque motion cues

We investigate the response of motion cells to stimuli representing homogeneous opaque motion patterns. We use (1) stimuli showing velocities matching the velocity tuning of the investigated cell and (2) stimuli with other velocities. We conduct experiments with varying}

---

4 The entire number of moving dots is 60% of the number of pixels in individual frames. If just one direction of motion is utilized (opaque motion) the density of the dots is 60%. If different directions are shown (transparent motion) the density of the dots may be less than 60% in individual frames caused by mutual occlusions of dots belonging to different movement planes.
Fig. 3. MT model cell responses to stimuli showing the preferred and anti-preferred direction (opposite velocities) of an observed cell with feature attention to the preferred and anti-preferred direction. Results are shown for six repetitions for each experimental configuration. The activity to the preferred direction is generally higher than to the anti-preferred direction. Note, that even if the anti-preferred direction is shown (black dots) attention to the preferred direction increases the neural activity compared to the case where the anti-preferred direction is attended.

configurations of attention, as described by Martinez-Trujillo and Treue (2004): (a) no attention, (b) attention to the velocity of the observed cell, and (c) attention to the presented velocity. Fig. 3 replicates the observation of Martinez-Trujillo and Treue (2004) showing that responses are higher when the preferred velocity is attended as compared to the case when the anti-preferred direction is attended. Fig. 4 shows this effect in more detail by investigating the activity of cells tuned to upward motion ($v_x = 0, v_y = 3$) for eight different presented motion directions (directions circularly arranged with fixed speed = 3). This experiment qualitatively replicates the observation of Martinez-Trujillo and Treue (2004) that feature attention to the preferred velocity increases neural activity in a modulatory fashion independent of the presented velocity, while attention to the presented velocity decreases the cells’ responses for velocities very dissimilar to the preferred velocity of the observed cell. Similar as in the study of Treue and coworkers the inhibitory effect depends on the dissimilarity of the presented velocity to the preferred velocity.

The results of both simulations described in this paragraph can be explained in terms of feedback modulation (modulatory excitation) coupled with lateral normalization. Particularly, the decrease of activity described in Fig. 4 by the dashed line (attention to the presented velocity, “attend same” in the experiment of Martinez-Trujillo and Treue) is a consequence of the normalization process described above. In combination with attentional modulation, this normalization implicates a decrease in activation of all other cells sharing the same spatial receptive field of the cell, which is highlighted by feature attention.

3.3. Influence of attention on the detectability of motion cues in transparent motion stimuli

Here, we investigate the ability to detect motion cues in transparent motion stimuli that consist of up to seven homogeneous layers of moving dots. The task of the model is to decide if a certain velocity was present in a given input sequence. We differentiate between two attentional situations: (a) no attention and (b) attention towards the velocity the decision has to be made. The major problem is the choice of an appropriate threshold applied to the neural activity in order to decide if a velocity is present or not. Thus, to generate results independent of the chosen threshold we apply a ROC analysis on the computed model output (Swets and Pickett, 1982). We use the numerically computed area under the ROC curves as a detectability indicator. This method basically tests all possible thresholds to describe the ability to decide if a velocity is existent in a presented stimulus. A higher value of this measurement

5 A random permutation determines the selected velocities (1–7 out of 8 directions with speed = 3 pixel/frame) for each input sequence (0 and 8 transparent layers were not presented because the answer to these stimuli is trivial). For each sequence eight runs are performed. In each run the model has to decide for one of the eight possible motions if it was present in the stimulus or not, such that all possible velocities are processed once.
Fig. 5. Detectability of individual velocities for different number of transparent layers (1–7) and different attentional conditions: attention to the preferred velocity of the observed cell (black dots) and no attention (white dots). To illustrate the detectability we show the area under ROC curve (auroc) based on the responses of the investigated cell for a set of 12 repetitions of each experimental condition (for each sequence eight different directions are investigated yielding $12 \times 8$ cell responses used for the calculation of the auroc): the maximum value for the auroc is 1.0 indicating a perfect discrimination with no false alarms, while a value of 0.5 represents random behavior. Left: auroc of the raw model MT cell responses. Right: auroc of the raw model MT cell responses with added Gaussian noise ($\sigma = 0.0002$; approx. 0.5% of the maximum activity over all experimental runs). The simulation demonstrates that (1) the detectability decreases with increasing number of transparent layers, and (2) attention to the preferred velocity increases the detectability and thus supports the decision process.

indicates a better detectability (1.0 is the maximum value and 0.5 represents random behavior).

Our simulations (Fig. 5) show that the knowledge about the decision to be made (feature attention) leads to higher detectabilities compared to cases without attention. In addition to the ROC analysis of the raw model output, we also added some noise to the neural activity of the investigated model MT cell to account for the decision process, which may be perturbed by noise terms. Our results are qualitatively consistent with the work of Felisberti and Zanker (2005), which however used absolute thresholds to evaluate the observer discrimination ability. Furthermore, the model cell activities are lowered in the presence of transparent motion pattern, which is consistent with Snowden et al. (1991) (data not shown). This effect is a simple consequence of mutual inhibition (normalization) between different model cells sharing the same spatial receptive field but tuned to different velocities. Our results concerning the detectabilities

![Graph](image1)

Fig. 5. Detectability for different number of feedback iterations using a stimuli with five layers of transparent motion. We show the area under ROC curve (auroc) based on the responses of the investigated cell with added noise for a set of 12 repetitions of each experimental condition as in Fig. 5. This result demonstrates that recurrent feedback processing (see model) increases the detectability in both cases, with and without attention (black and white dots, respectively). Note, particularly in the case without attention the results seem to vary over time. This is explained by the high number of presented transparent layers which induce high ambiguities and which cannot be resolved. Those competing activities lead to fluctuations in the discriminability over time.

![Graph](image2)

For each run a set of 100 values is computed by adding Gaussian noise to the computed neural activity with a standard deviation of approx. 0.5% of the maximum value over all experiments (see Fig. 5). The choice of noise level was arbitrary and (much) stronger noise may lead to undesired effects such that the velocity cannot be extracted any more.
(Fig. 5) can be explained by the increased neural activity coupled with the additive noise of constant strength leading to a better signal-to-noise ratio in the presence of attention. However, even without additional noise we observe an increased detection performance suggesting that the effects of recurrent iterative feedback processing lead to motion signals that were easier to detect than with a simple multiplicative mechanism.\footnote{A pure multiplication of a cell’s activity with a (constant) attention signal does not change the ability to detect a cue based on this signal. In other words, the relative overlap of two distributions representing “cue present” and “no cue present” remains unchanged, irrespectively of a factor that is multiplied on the investigated value.} Fig. 6 substantiates this claim by showing that the detectability increases over time during feedback processing.

4. Conclusion

We utilized a model of motion perception to explain physiological data concerning feature attention as well as data from psychophysical studies.

We qualitatively reproduce physiological results (Martinez-Trujillo and Treue, 2004) which suggest an excitatory modulation of cell activities if the preferred motion is attended and inhibitory interaction if the anti-preferred motion is attended. Our model explains these effects in terms of attentional modulation and subsequent mutual competition in a recurrent framework of biased competition.

The pure modulation suggested by Martinez-Trujillo and Treue (2004) would not result in differences in psychophysical motion detection tasks, unless (1) other than pure modulatory effects are involved or (2) noise is added to the investigated neural activity. Our model is able to explain results of a psychophysical study of motion detection (Felisberti and Zanker, 2005) (1) by its iterative recurrent processing scheme, which implies a more complex integration than a pure modulation of the output and (2) by adding noise to the output prior to the decision process. Using ROC analysis we demonstrate how iterative recurrent information processing further increases the detectability of certain velocities. In addition, we illustrate how iterative recurrent information processing further increases the detectability.

By combining the results obtained by our model with respect to both investigated experimental studies we make a psychophysically testable prediction. Our investigations suggest a decrease of the detectability in the experiment of Felisberti and Zanker (2005) if attention is directed towards another velocity than the velocity of interest. This claim is deduced from the fact that neural activity is decreased if a velocity very different to the preferred velocity of the investigated cells is attended. Such lowered activity level in turn leads to an increased signal-to-noise ratio in the following decision process assuming noise independent of the outcome of area MT.

In conclusion, our model brings together physiological and psychological data and gives deeper understanding of possible mechanisms of motion processing, the role of neural activities at the stage of MT for motion detection tasks, and the influence of feature attention in visual motion tasks.

Acknowledgments

This research has been supported in part by a grant from the European Union to Heiko Neumann (EU FP6 IST; project number 027198).

Appendix A. Model equations

Here we present the equations and the corresponding parameterizations used for all simulations in this paper (compare Bayerl and Neumann, 2004).

The input stage of model V1 computes normalized oriented contrast responses \( c^{(1)} \) (Eq. (1); eight orientations specified by \( \alpha \) and \( \beta \)), raw correlation signals corresponding to half-detectors \( c^{(2)} \) (Eq. (2); compare Hassenstein and Reichardt, 1956), and a motion signal \( c^{(3)} \) (Eq. (3)) corresponding to a full Reichardt detector with divisive inhibition (Hassenstein and Reichardt, 1956; Bayerl and Neumann, 2004). Note, that for the normalization in Eq. (1) a small constant of 0.01 is employed to prevent divisions by zero:

\[
  c^{(1)} = \frac{\left[ I \ast \frac{\partial^2 c^{(1)}(x, \alpha, \beta)}{\partial x^2} G_{\sigma = \sqrt{2}} \right]_+}{0.01 + \sum_\beta \left[ \frac{I \ast \frac{\partial^2 c^{(1)}(x, \alpha, \beta)}{\partial x^2} G_{\sigma = \sqrt{2}} \right]_+ * G_{\sigma = 1.0}}
\]  

\[
  c^{(2+)} = \left( \sum_\alpha c^{(1)}(x, \Delta x, t + 1) \right) * G_{\sigma = 1.0},
\]

\[
  c^{(2-)} = \left( \sum_\alpha c^{(1)}(x, \Delta x, t + 1) \right) * G_{\sigma = 1.0}
\]

\[
  c^{(3)} = \frac{c^{(2+)} - c^{(2-)}}{1.0}
\]

where \( c^{(1)} \) is a function of space and orientation and \( c^{(2)} \) and \( c^{(3)} \) are functions of space and velocity. \( \ast G \) denotes the convolution operation, \([\cdot]_+ \) is the rectification operation.
operation \( \max(-,0) \), \( \sigma^2 \) the second directional derivative in direction \( \alpha \), and \( G_{\sigma} \) a spatial Gaussian with standard deviation \( \sigma \). \( x \) denotes the spatial location (2D), \( \Delta x \) the spatial shift per frame corresponding to a velocity, and \( t \) the time (frame number of the input sequence).

The dynamics within model areas V1 and MT is described by the following equations (Eq. (4–6)): \( v^{(1)} \) modulates the input signal with feedback (or attentional) modulation (Eq. (4)). This operation enhances patterns in the input signal compatible to the expected signal given by the feedback or attentional signal. Note, that for early feedback between V1 and MT top-down signal connections are shifted according to the encoded velocity of each cell to implement the tracking of moving patterns (e.g., a model MT cell tuned to 3 pixels to the left is linked by a feedback connection to a model V1 cell with a receptive field located 3 pixels to the left of the MT cell). \( v^{(2)} \) realizes the feedforward integration of the modulated input signal (Eq. (5)): the nonlinearity (square operation) enhances differences between competing activities and, thus, speeds up the relaxation process. Spatial integration combines neighboring motion cues in model MT and integration in velocity space allows combining similar velocities and stabilizes the output of both model areas. \( v^{(3)} \) implements a local shunting inhibition to normalize the output of the model by keeping the local sum of activities in certain bounds (Eq. (6)). A consequence of this normalization is that an increase of activity for any velocity in turn generates a decrease of all other activities sharing the same spatial location.

\[
v^{(1)} = (\text{net}_{\text{IN}} + A) \cdot (1 + C \cdot \text{net}_{\text{FB}}) \quad (4)
\]

\[
v^{(2)} = (v^{(1)})^2 \ast G^{(\text{space})} \ast G^{(\text{velocity})} \quad (5)
\]

\[
v^{(3)} = \frac{v^{(2)}}{B + \sum_{\text{all velocities}} v^{(2)}} \quad (6)
\]

where \( v^{(1)} \), \( v^{(2)} \), \( v^{(3)} \), \( \text{net}_{\text{IN}} \) and \( \text{net}_{\text{FB}} \) are functions of space and velocity. \( G \) denotes Gaussian kernels in spatial and velocity domains.

The parameterization of the model used for all simulations in this paper is given in Table 1.

The attentional signal \( a \) is zero in experiments without attention. Else it is a Gaussian in velocity space with \( \sigma = 2.0 \) with a peak activity of 0.2 centered at the attended velocity (independent of the location).

The model activities are computed iteratively until the model relaxes (after 4–6 iterations in general). This is accomplished by (re)computing Eqs. (1)–(6) in the presented order at each time step (Eqs. (4)–(6) are computed twice, once for each of both model areas). If not mentioned otherwise (Fig. 6) we show the relaxed model state after eight iterations of feedback processing. In all figures in the paper we evaluate the output of model area MT (\( v^{(3)} \)). In the example shown in Fig. 2 the maximum value of this variable at each location is evaluated and the velocity corresponding to the cell with the maximum activity is interpreted as detected velocity. In all other figures, response of the model with respect to an individual velocity is computed by sampling the activity of the cell tuned to this velocity with a receptive field located at the center of the visual field.

References


