Perceptual Motion Illusions as a Tool to Probe Neural Mechanisms of Motion Integration in the V1-MT-MSTl Feedforward-Feedback System

Daniel Schmid (daniel-1.schmid@uni-ulm.de)
Institute of Neural Information Processing, Ulm University
James-Franck-Ring, 89081 Ulm, Germany

Maximilian P. R. Löhr (maximilian.loehr@uni-ulm.de)
Institute of Neural Information Processing, Ulm University
James-Franck-Ring, 89081 Ulm, Germany

Heiko Neumann (heiko.neumann@uni-ulm.de)
Institute of Neural Information Processing, Ulm University
James-Franck-Ring, 89081 Ulm, Germany

Abstract

Visual motion integration needs to resolve ambiguous or conflicting information. While for most stimuli subjects are able to perceive the stimulus motion correctly, subjects fail to do so for motion illusion stimuli. In this work we use such illusion stimuli, namely drifting Gabor wavelets, to probe a hierarchical computational neural model of V1-MT-MSTl for its mechanisms of motion integration via recurrent feedforward-feedback interactions. We find that later stages are more susceptible to illusory motion, while earlier stages closely capture the true stimulus location. By lesioning of feedback connections we show that the effect can be explained by feedforward computation alone already. We conclude that cortical top-down feedback within the employed model serves as a predictive element besides taking part in linking information across neural model columns.

Keywords: neural motion integration; spiking model; area MT; cortical feedback; motion illusion

Introduction

Our visual system selectively integrates multiple related inputs and segregates them from unrelated ones. In the case of motion input stimuli are often ambiguous (aperture problem) and even localized features need to be evaluated and distinguished according to their intrinsic or extrinsic surface belongingness. Since cells in area V1 only have small receptive fields (RF), they only sense input components of grating patterns, while area MT cells have larger RF and selectively combine multiple responses with disparate velocity attributes (Adelson & Movshon, 1982).

An open question remains how component inputs are integrated, particularly at the stage of MT. Different stimulus configurations have been used in various experimental investigations. For example, oriented bars of different lengths have been used to evaluate the time to disambiguate the aperture problem (Pack & Born, 2001; Born, Pack, & Zhao, 2002) while other studies used plaids synthesized from gratings of different velocity compositions to evaluate motion integration mechanisms such as vector average or intersection-of-constraints

(Adelson & Bergen, 1985; Welch, 1989). More recent investigations used multiple oriented bars presented apart or overlaid to justify the selectivity of integration in area MT cells as pattern- and component-direction selective (Smith, Majaj, & Movshon, 2005; Majaj, Carandini, & Movshon, 2007).

Neural models have been proposed to explain the mechanisms of motion composition in the V1-MT-MST cascade. Such models can be categorized as selectionist or integrationist to account for the input features they utilize for motion integration and disambiguation (Pack & Born, 2008). Models can be further distinguished into feedforward (Adelson & Bergen, 1985; Rust, Mante, Simoncelli, & Movshon, 2006) and recurrent feedforward-feedback approaches (Grossberg, Mingolla, & Viswanathan, 2001; Bayerl & Neumann, 2004; Tlapale, Masson, & Kornprobst, 2010). More recently, we have proposed a model architecture composed of spiking neurons for hierarchical motion analysis building upon detailed findings about the response characteristics of V1, MT, and MSTl cells and their recurrent and bottom-up and top-down feedback interactions (Löhr, Schmid, & Neumann, 2019a). We argue that the details of how component stimuli are integrated into coherent motion percepts can be revealed in part by displays which contain conflicting local input motion configurations. Freeman, Adelson, and Heeger (1991) showed that local modulation of spatio-temporal phase of complex Gabor filters leads to apparent movement percepts. Variants of such patterns have been utilized to experimentally probe visual motion perception (Tse & Hsieh, 2006).

In this work we make use of stimuli, that induce motion without movement, as well as their composition into apparently moving shapes. In particular we employ the curveball illusion (Shapiro, Lu, Huang, Knight, & Ennis, 2010) to probe the relative influence of stationary and non-stationary movement components (cf. Figure 1). The apparent inward/outward motion (Whitney et al., 2003) is used to demonstrate the integration of disparate movement evidences (cf. Figure 1). Such probing of the neural model architecture for motion integration (Löhr et al., 2019a) helps us to separate the contributions of different model components to feedforward motion integration and predictive integration along the feedback projections.
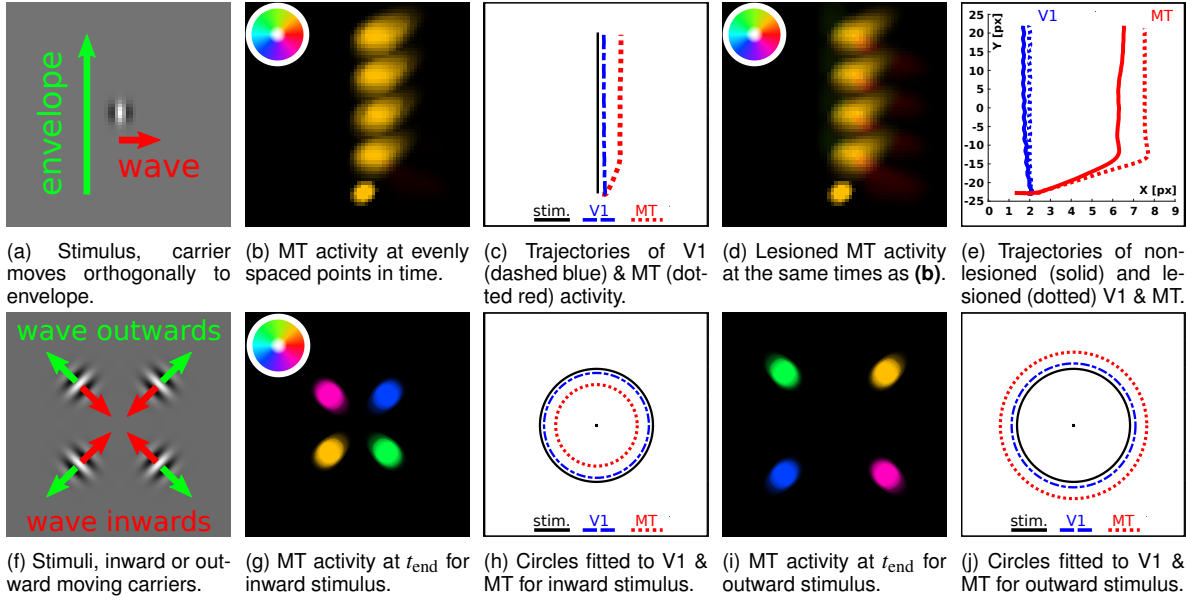
(a) Stimulus, carrier moves orthogonally to envelope.

(b) MT activity at evenly spaced points in time.

(c) Trajectories of V1 (dashed blue) & MT (dotted red) activity.

(d) Lesioned MT activity at the same times as **(b)**.

(e) Trajectories of non-lesioned (solid) and lesioned (dotted) V1 & MT.

(f) Stimuli, inward or outward moving carriers.

(g) MT activity at t_{end} for inward stimulus.

(h) Circles fitted to V1 & MT for inward stimulus.

(i) MT activity at t_{end} for outward stimulus.

(j) Circles fitted to V1 & MT for outward stimulus.

Figure 1: **(a-e) Stimulus and results of experiment I. (a)** Stimulus is a Gabor patch ($\sigma = 1.83$ px, $\omega = \pi/3$ px). Its envelope moves vertically at 45 px/s and its carrier drifts horizontally at 45 px/s. **(b)** The spatial distribution of neuronal activity of MT at five points spaced evenly in time. **(d)** Same as **(b)**, but for the lesioned MT area. Color encodes direction of motion (inset) and black denotes zero activity. **(c)** Trajectories of stimulus (solid black) and V1 (dashed blue) and MT (dotted red) neuronal activity. **(e)** The effect of lesioning. Trajectories for the lesioned model (dotted) show a higher susceptibility for the drift component as compared to the unlesioned model (solid). **(f-j) Stimuli and results of experiment II. (f)** Stimulus consists of four Gabor patches ($\sigma = 6.83$ px, $\omega = \pi/6$ px) on a circle of 40 px radius. Envelopes are stationary but all carriers either drift inwards or outwards at 64 px/s. **(g,i)** Spatial distribution of neuronal activity of MT once for the inward and once for the outward stimulus at the end of simulation. **(h,j)** Circles fitted to the stimulus (solid black) and to neuronal activity of V1 (dashed blue) and of MT (dotted red) for both stimuli. At the end of simulation these fitted radii reach $r_{\mathrm{inw.}}^{\mathrm{V1}} = 37.1$ px and $r_{\mathrm{inw.}}^{\mathrm{MT}} = 28.8$ px for the inward and $r_{\mathrm{out.}}^{\mathrm{V1}} = 43.8$ px and $r_{\mathrm{out.}}^{\mathrm{MT}} = 51.8$ px for the outward stimulus. **(a-d)** span the same coordinate space and so do **(f-j)**.
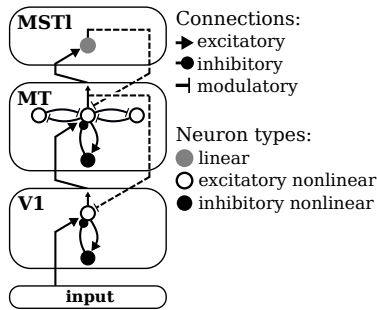

Figure 2: Neural model architecture of the visual cortex used in our investigations. See Methods for details.

Methods

Neural Architecture of Visual Motion Integration

The model proposed by Löhr, Schmid, and Neumann (2019b); Löhr et al. (2019a) consists of a hierarchy of areas V1, MT and MSTl (Figure 2). Visual input to V1 is fed into neurons either selective for static or moving oriented contrasts. At this stage static and moving representations pose conflicting hypothe-

ses that suppress each other at the same spatial position. V1 oriented motion information pooled and smoothed over a local neighborhood is fed into MT, where neurons are selective for different speeds per orientation. MT's different motion hypotheses take into account a lateral neighborhood and mutually compete via pooling to reach a consolidated state. The consolidated representation of MT acts upon V1 via feedback connections gradually guiding it, without completely supressing the original locally available information. Spatial integration over local neighborhoods of V1 via MT and of MT via MSTl and feeding back this integrated information are responsible for spreading hypotheses accross spatial positions, but only to neural sites where activity was present originally as well (Löhr et al., 2019b, 2019a).

Each area is represented as hypercolumns at every spatial position of the input. Hypercolumns are modeled by pairs of excitatory-inhibitory spiking neurons. The coupling between these can either be excitatory, inhibitory, or modulatory. The interaction of these different types of input to a model neuron happens by a dynamical process, described as ordinary differential equations (ODEs), as formerly investigated by the canonical neural circuits model (CNCM (Brosch & Neumann,

2014)). The CNCM is a conductance-based model, that modulates feedforward excitatory input into excitatory cells by a feedback signal and normalizes columnar responses via a pooling mechanism established by inhibitory cells. This interaction principle allows bottom-up information to be upregulated by feedback. Feedback stems from higher order model areas (top-down modulation), or in the case of MT additionally from other spatial positions of the same area (lateral modulation). So modulation together with inhibitory pooling leads to a consolidation of the hypercolumnar representation. The connectivity between neurons is modeled by weighting kernels (in spatial, temporal and feature dimensions). Interactions are realized as filtering operations using these kernels. The model is implemented in MATLAB 2018a and an Euler scheme is used to solve the ODEs. The parameterization and further details on the model can be found in Löhr et al. (2019a).

Experiments

The experiments are inspired by psychophysical experiments using drifting Gabor stimuli. In the model increased activity codes for more important information. Model results are evaluated w.r.t (a) the readout of the population code of area MT for the angle of the encoded motion to inspect how available motion information is taken into account and (b) the position of the center of mass of the population code of areas V1 and MT to check for possible changes in spatial processing of the stimulus along the visual hierarchy.

I - Curveball Illusion The first set of experiments is based on the curveball stimulus Shapiro et al. (2010). The model is shown a Gabor stimulus moving from the bottom to the top while its carrier drifts from left to right (cf. Figure 1). Besides the full model architecture a version with lesioned feedback connections is investigated (cf. dashed connections in Figure 2), where all top-down feedback connections are cut.

II - Apparent Inward/Outward Motion For the second set of experiments a stimulus is used similar to the one used by Whitney et al. (2003). Responses are computed for the full model on an input consisting of four stationary Gabor patches, that are arranged on a circle (Figure 1). Two conditions are simulated. One where all carriers are drifting inward and one where all carriers are drifting outward w.r.t the center.

Results

I - Curveball Illusion

The center of mass of MT's population response shifts gradually over time from an initial position close the stimulus towards one offset into the direction of the drift (Figure 1). This is in accordance with psychophysical results from Shapiro et al. (2010) in the case of peripheral stimulus representation. Comparing responses of V1 and MT it becomes apparent, that the positional displacement is much more pronounced in MT than in V1. Moving up the visual model hierarchy the response characteristic shifts from a more input-based one in V1 to a more percept-based one in MT.

The overall orientation hypothesis of MT's population response is tuned to a superposition of the motion vectors from the envelope's motion and the drift vector of the Gabor (Figure 1). Thus, the model shows the same integrationist behavior as on previously tested stimuli by Löhr et al. (2019b, 2019a). Here a common underlying mechanism is able to integrate localized motion information, not just extracted across space from oriented contours (in the case of Löhr et al. (2019b, 2019a)), but also extracted across different components aligned in space. The model's ability to do so stems from its hierarchical architecture first extracting simple feature information (common to oriented contrasts, drifting Gabor carriers and moving Gabor envelopes) and integrating and consolidating it afterwards at a higher stage.

When lesioning V1 and MT their trajectories shift away more from the curveball's trajectory. Concerning the main findings of the unlesioned model, that the higher stage MT closer resembles the perceived, while V1 closer resembles the true trajectory, hold as well for the lesioned case. Interestingly, while MT's response is biased towards perceived motion, its feedback shifts V1's activity closer towards the true trajectory (Figure 1). Thus, the main finding can be explained by missing feedback, that formerly projected back to lower levels of the hierarchy at extrapolated positions, thus making predictions of how the stimulus will move. In the full model this prediction matched the envelope movement, while mismatched the stationary drift component, since it didn't move to the extrapolated position, guiding the representation towards the envelope motion.

The lesion, thus, shows that the motion-induced position shift can already be explained by feedforward processing of the hierarchy alone without the need for feedback. We argue that the main property forming the different displacements in V1 and MT are the RF properties of either stage. While V1 has smaller kernels, that act directly upon the input, MT has spatio-temporally more elongated kernels acting upon V1 responses. This stronger spatio-temporal elongation of model MT's kernels leads to an increased responsiveness of the MT RF displaced in space and time. Since V1's dynamical processes give prolonged activities across the most recent positions which the curveball has passed, MT's kernels respond best if they are integrating with the respective displacement their tuning dictates. This can be interpreted in such a way, that MT's kernels respond best at that position their tunings hypothesize the motion to occur in the near future, thus extrapolating from the recent input.

II - Apparent Inward/Outward Motion

Responses of V1 and MT at every patch location in the second experiment are shifted along the drift direction of the stationary Gabor patches (Figure 1). This resembles results from psychophysical investigations of Whitney et al. (2003), their Figure 1. Again, a much more pronounced shift is found in MT's response, while the response of V1 more closely resembles the true positions. This validates the findings concerning the first experiment, while extending it to purely stationary

configurations at the same time.

Conclusion

In the present work we showed how illusory motion stimuli can be used to investigate hierarchical feedforward-feedback processes in a computational model hierarchy of V1-MT-MSTl. The performed simulations give new insights into the role of the processing stages and their feedback connections. Lesion studies of these connections reveal that the main findings of our experiments can be explained by hierarchical feedforward processing within the model. These findings together with the discoveries of Löhr et al. (2019b, 2019a) show that feedback most prominently serves to integrate and form a coherent representation across spatially distributed information as well as aligning activity by propagating a predictive representation to earlier levels of the hierarchy.

The model was able to integrate information of curveball's different motion components, as well as of several disparate stationary drifting Gabors. Further, we identified a stronger resemblance of the true input stimulus motion by area V1 and resemblance more closely related to illusiory motions and implied shapes by area MT. The RF properties of MT led to a response profile spatially extrapolating the detected motion. The direction tuning of MT's response showed clear integration of different motion components of the stimuli giving exemplary insight into how integrationist models are processing illusory motion stimuli. This extends the findings of Löhr et al. (2019b, 2019a) from sets of spatially distributed motion information of moving oriented contrasts.

Model MT's response is torn towards illusory motion by drifting Gabors. The model encompasses the extraction and consolidation of available motion information by the visual cortex, but doesn't model the complete brain. This leaves ample space for higher order processes to act upon this representation, e.g. to alter percepts in foveal regions (Shapiro et al., 2010) or to reorganize the activity distribution towards the counterdirection of a drifting arrangement (Whitney et al., 2003), and hence invites to further investigation.

Acknowledgments

References

Adelson, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A*, *2*(2), 284.

Adelson, E. H., & Movshon, J. A. (1982). Phenomenal coherence of moving visual patterns. *Nature*, *300*, 3.

Bayerl, P., & Neumann, H. (2004). Disambiguating visual motion through contextual feedback modulation. *Neural Computation*, *16*(10), 2041–2066.

Born, R. T., Pack, C. C., & Zhao, R. (2002). Integration of motion cues for the initiation of smooth pursuit eye movements. In *Progress in Brain Research* (Vol. 140, pp. 225–237). Elsevier.

Brosch, T., & Neumann, H. (2014). Computing with a canonical neural circuits model with pool normalization and modulating feedback. *Neural Computation*, *26*(12), 2735–2789.

Freeman, W. T., Adelson, E. H., & Heeger, D. J. (1991). Motion without movement. In *Proceedings of the 18th Annual Conference on Computer Graphics and Interactive Techniques* (pp. 27–30). New York, NY, USA: ACM.

Grossberg, S., Mingolla, E., & Viswanathan, L. (2001). Neural dynamics of motion integration and segmentation within and across apertures. *Vision Research*, *41*(19), 2521–2553.

Löhr, M. P. R., Schmid, D., & Neumann, H. (2019a). Motion integration and disambiguation by spiking V1-MT-MSTl feedforward-feedback interaction. In *2019 IJCNN* (Vol. in press).

Löhr, M. P. R., Schmid, D., & Neumann, H. (2019b). Motion integration and disambiguation concerted by feedforward-feedback interactions of V1-MT-MSTl. In *Journal of Vision* (Vol. in press).

Majaj, N. J., Carandini, M., & Movshon, J. A. (2007). Motion integration by neurons in macaque MT is local, not global. *Journal of Neuroscience*, *27*(2), 366–370.

Pack, C. C., & Born, R. T. (2001). Temporal dynamics of a neural solution to the aperture problem in visual area MT of macaque brain. *Nature*, *409*(6823), 1040–1042.

Pack, C. C., & Born, R. T. (2008). 2.11 - Cortical mechanisms for the integration of visual motion. In R. H. Masland et al. (Eds.), *The Senses: A Comprehensive Reference*. New York: Academic Press.

Rust, N. C., Mante, V., Simoncelli, E. P., & Movshon, J. A. (2006). How MT cells analyze the motion of visual patterns. *Nature Neuroscience*, *9*(11), 1421–1431.

Shapiro, A., Lu, Z.-L., Huang, C.-B., Knight, E., & Ennis, R. (2010). Transitions between central and peripheral vision create spatial/temporal distortions: A hypothesis concerning the perceived break of the Curveball. *PLoS ONE*, *5*(10), e13296.

Smith, M. A., Majaj, N. J., & Movshon, J. A. (2005). Dynamics of motion signaling by neurons in macaque area MT. *Nature Neuroscience*, *8*(2), 220–228.

Tlapale, E., Masson, G. S., & Kornprobst, P. (2010). Modelling the dynamics of motion integration with a new luminance-gated diffusion mechanism. *Vision Research*, *50*(17), 1676–1692.

Tse, P., & Hsieh, P.-J. (2006). The infinite regress illusion reveals faulty integration of local and global motion signals. *Vision Research*, *46*(22), 3881–3885.

Welch, L. (1989). The perception of moving plaids reveals two motion-processing stages. *Nature*, *337*(6209), 734–736.

Whitney, D., Goltz, H. C., Thomas, C. G., Gati, J. S., Menon, R. S., & Goodale, M. A. (2003). Flexible retinotopy: Motion-dependent position coding in the visual cortex. *Science*, *302*(5646), 878–881.