Sean Lorenz

CN530

12/12/2007

CN530 Final Report

Abstract

How a moving input stimulus becomes a direction percept via processing in the human visual system is a complex and unresolved issue which involves numerous areas of the brain. One of the principal issues in determining direction percept entails decoding of the numerous signals are honing these signals down into one percept. More specifically, how do numerous orientation, disparity, and speed signals from V1 and V2 reduce to form a motion percept in MT, presuming MT is the pooling locus of motion activity. I then give four strategies as an overview in this field for explaining various motion percept tasks: winner-take-all, vector averaging, Bayesian, and a sigmoid feedback function. I conclude by elaborating upon a model by Van Wezel and Britten which details several of these percept strategies, offers a downstream mechanism to account for behavioral visuomotor outputs, and may also support the idea of a sigmoid feedback function being implemented in direction percept and readout of MT.

Introduction

The ability to detect motion in the human visual system is an evolutionary boon, enhancing our species' ability to evade predators and, in turn, visually track moving prey when hunting. So in what evolved brain area, or areas, do we code for movement across a visual scene with complex textures and objects moving in numerous directions at different depths? The middle temporal visual area (MT), which is also called V5, has been studied extensively over the past several decades and has since proven to be a key component of motion perception

processing. Located directly in the middle of the dorsal visual stream, MT receives strong V1 magnocellular inputs as well as projections from V2 thick stripes, V3, and inferior pulvinar.

Like V1, MT is retinotopically organized in each hemisphere, however, MT cells have far larger receptive fields which code for preferred direction as well as binocular disparity and speed in a columnar structure (Born & Bradley, 2005). V1 already codes for many of these features, however, due to the large MT receptive fields, MT neurons can account for motion and disparity in a much larger spatial domain. More importantly, MT is considered by many researchers to be the locus of visual motion perception and also may be considered a branching point for determining behavioral motor outputs such as object tracking or saccadic generation. It is obvious from these feature qualities that MT is more concerned with decoding motion and spatial location than object recognition or color.

Problems With Direction Percept

In order to obtain a direction percept, MT must compute the velocity of input taken in from the retina, filter out and suppress noise signals sent from V1, and pool or extract a meaningful motion signal from this noisy system in order to segment what is moving or "figure" from what is static or "ground" in a visual environment. All this processing may be why MT neurons are sensitive to not only preferred direction, but also overlapping features including disparity and speed. So how do numerous sensory signals become a solitary perceptual decision of direction in MT? And how does this "readout" effect visuomotor behavior?

Much of the MT literature describes this area's neural response as having a Gaussian bell shape as seen in a simple motion psychophysical test where a white dot moves across a black screen in one direction. This Gaussian distribution correlates to numerous synaptic inputs from areas such as V1 across MT neurons' large receptive fields, which must then be reduced to a

2

single percept, signal, or code. In order to determine one percept, a method of choice from an array of activity must be taken within MT. How this choice is made has been of particular interest to vision researchers and has spawned numerous models attempting to interpret direction percept selection and visuomotor behavioral output.

Based on the limited number of papers I have read on this topic thus far, it appears that two primary methods have been proposed: 1) the *most active* neuron(s) in the input network "win" the competition and hence become the motion percept or 2) the *mean activity* across the network is computed and a certain spatiotemporal distribution, or "vector averaging," determines the percept. Unfortunately, neither of these two methods on their own have been able to fully account for MT responses over a wide range of psychophysical motion percept experiments. It appears that MT uses the first, winner-take-all strategy on some occasions and the second, vector averaging strategy on other occasions. Born and Bradley mention a third option for decoding the motion percept – a model which relies on Bayesian methods. I will discuss the basic tenets of this model type in a later section as well as some issues with a Bayesian approach to direction percept. In turn, I will consider a fourth option for decoding MT direction percepts by implementing a sigmoid feedback function.

Winner-Take-All

The Newsome Lab at Stanford is a chief contributor to discussion of sensory map read out with special emphasis on MT and its relation to visuomotor output at this level. In other words, they believe MT in integral to production of a proper motion signal for visually guided motor commands such as saccades or smooth pursuit eye movements. Groh and colleagues have also performed several electrophysiology experiments on rhesus monkeys in which visual target stimuli (**V**) were presented for fixation, saccade, and combined saccade/pursuit tasks (Groh et al.,

1997). In addition to this visual stimulus, an artificial velocity signal (**E**) produced by microstimulation of an MT column was introduced while the monkey either fixated on a point, made a saccade to moving target, or tracked a moving target.

A winner-take-all algorithm would predict that the winning percept (either the **V** or **E** vector) depends on which signal wins the competition. So, if the **V** vector were to win, the monkey should see the normal visual stimulus and ignore the electrode stimulation, whereas if the **E** vector were to win the opposite would be true. In other words, this method picks the more salient of two stimuli choices. Were both **V** and **E** of equal weights or activity in this set up, one would expect the winner to be probabilistic, varying from trial to trial.

The results of this experiment demonstrates that the pattern across target velocities in MT used different mechanisms for different tasks. It was believed that a winner-take-all schema was chosen to segregate distinct motion signals and may be most useful when more than one moving object or pattern overlaps in the visual space. For the most part, however, this algorithmic method for defining the motion percept was not seen often in their experiments. In a later experiment by the same lab (Nichols & Newsome, 2002) it was shown that, in perceptual tasks, neurons in MT with preferred direction differences *over* 140° competed for direction percept dominance. To account for preferred direction differences *less than* 140°, it was hypothesized that disparate motion signals cooperated in order to influence the direction estimate. This brings us to the next topic of vector averaging.

Vector Averaging

In the experimental representation mentioned above, the researchers found that the pattern of stimulation-induced errors in the saccade and pursuit tasks gave evidence that a vector

averaging algorithm was used to decode the bell-shaped MT input in order to produce a single motion vector percept. Vector averaging is the notion that preferred velocities are averaged together and scaled according to their relative input strength. This strategy may be important for quick decisions in saccade generation since it can rapidly approximate motion information in small spatial regions in order to track or fixate on an important region of the visual scene. Such a strategy would seem to have extraordinary evolutionary significance which is why it may be a common algorithmic method used by MT.

It is important to note that the outcome of vector averaging always reflects the presence of *all stimuli* entering MT, which is the opposite scenario from a winner-take-all setup. There have been numerous experiments performed, most of which seem to agree with the idea that vector averaging is the method of choice for initial direction of smooth pursuit, saccadic velocity compensation, and choice of motion percept. It is pointed out that errors to averaging predictions for short-duration trials were much smaller than winner-take-all predictions whereas the opposite is true for long-duration trials (van Wezel & Britten, 2002). Also, MT cells seem to be switching from an averaging algorithm to a winner-take-all preference after around 200ms – the time at which pursuit movement tends to take place. Thus, it seems that fast readout relies on vector averaging in MT yet this mechanism switches over to winner-take-all once tracking of an object becomes necessary. This 200ms latency, however, may be due to other areas interacting with MT in downstream areas such as MST or posterior parietal cortex (PPC).

As shown above, vector averaging relies on combining direction signals in order to formulate a single motion percept, yet it seems to work best for only *simple* stimuli. It may be for this reason that such an algorithmic method cannot fully explain more complex illusory stimuli like the moving plaids. It would be also important to know how MT accounts for multiple

objects moving in various directions for every day scenarios in addition to psychophysical experiments controlled in a lab. A recent proposal to solve this problem relies upon a Bayesian model of motion perception.

Bayesian

So far I have proposed that the direction percept correlates to discrete activity peak choice or activity pooling, yet another element should be included in this picture – the notion of uncertainty. It could be said that integration of V1 and LGN information into more global motion descriptions is an ambiguous measurement procedure, thus one can account for a certain amount of uncertainty when speaking of motion percepts (Weiss et al., 2002). Weiss mentions that intersection of constraints, vector averaging, and feature tracking together still do not give a holistic model for predicting perceived direction velocity.

The solution proposed is based on Helmholtz's view of percepts being a best guess of what is going on in the world coupled with a Bayesian estimation framework. The model is based on two assumptions: 1) local image measurements are noisy and 2) image velocities tend to be slow. The velocity estimate is calculated from one equation which takes the mean or maximum of the posterior Gaussian distribution and predicts an ideal observer's velocity estimate for any image sequence. This single Bayesian algorithm was able to correctly forecast a number of psychophysical results, such as the plaids, once a nonlinear gain control function was later added to the model. The authors state that the abstracted perceptual noise model presented here can be seen as a way of accounting for the uncertainties of neural responses in MT.

The Bayesian concept described above accounts for noise effects and prior weight biases yet cannot explain complicated motions such as rotations or even multiple motions in general. The larger issue here, however, is that a Bayesian approach is able to produce a certain

perceptual outcome by tweaking certain algorithmic parameters, yet completely bypasses the fact that actual neurons are wired together and synaptically functioning in specific networks within the brain. Instead, Weiss's Bayesian method acts like a revived behaviorism mathematical model which merely looks at an input and produces a workable algorithm for perceptual output. There is indeed some element of noise and chance in neural activity, yet a purely probabilistic model of biological MT neural function (even with some weighting factors involved) seems untenable.

Sigmoid Function

The Newsome Lab research states that evidence for each of the direction percept algorithms is due to variance on the kind of task being performed, yet this may not be the only explanation. In order to know the underpinnings of the visual system's motion networks, it seems as if there may be some degree of task invariance which explains transition in algorithm usage for MT neurons. Since there are times when these signals behave in a winner-take-all manner and in an averaging manner elsewhere, it may be possible to explain this algorithmic variation using a sigmoid feedback function instead (Grossberg, 1973).

This fourth option is not a strictly bottom-up method as are the prior three concepts, thus it adds another level of complexity. The three major properties of this function are: 1) its ability to contrast enhance small signals, 2) store intermediate signals with little distortion, and 3) uniformize very large signals. The sigmoid signal itself is made of an initial faster-than-linear function which acts as a winner-take-all pattern storage and "quenches" small, noisy activities. The middle section of the function is linear which means that after the small activities are quenched a perfect pattern is stored. Finally, the last portion of the function is slower-than-linear, and thus it reduces contrast and increases uniformization. How much of a network's activity is quenched can be adjusted in order to determine the resulting pattern.

This type of method relies on a shunting on-center off-surround network with feedback, which means that for MT neurons the output of the population would be a sigmoid function of its activity. Interestingly, the weighted average of the feedback gains from a sigmoid feedback functions exhibits the same bell-shaped curve as the MT neuron distribution. These last few statements are purely speculative and give rise to a host of other questions, namely asking where (if at all) the recurrence and feedback stem from in order to adequately account for a variety of motion percept tasks.

A Few More Thoughts on Sensory Readout

Van Wezel and Britten believe adaptation effects occur downstream of MT where a winner-take-all step is more likely to be taken. At the level of MT, effects of adaptation are very selective, selectively attenuating the adaptation direction and leaving the opposite direction alone. Thus adaptation effects become symmetric with stimulus direction and may be a loose parallel to the notion of a quenching threshold via recurrence from downstream activity.

So how does this all relate to the problem of readout from sensory maps to cortical maps in order to generate percepts or motor behaviors? Van Wezel & Britten propose that different downstream structures have different rule sets that run in parallel on the output from sensory maps such as the one for MT. Returning to an earlier statement about MT cells switching from winner-take-all to averaging behavior, the latency seen here may be due to downstream competitive mechanisms which could implement a Grossbergian sort of sigmoid feedback function. These same authors propose a model with three linked parts: 1) sensory maps in cortex are task invariant, 2) basic rules for varying combined signals are also relatively invariant and thus readout rules remain somewhat stable, and 3) multiple readout mechanisms are able to coexist for varying tasks. One thing that is missing from this model is a notion of synaptic

plasticity and how this would effect the sensory maps and the notion of direction percept over time.

In this type of model, vector averaging can take place within the MT while faster-than-linear decisions are made in parietal or frontal cortex regions. Variation in tasks is seen as accessing different downstream circuits in parallel, which may also be feeding back to MT and MST. This setup attempts to account for an MT direction percept as well as describing how signals to premotor areas could be produced for behavioral outputs such as saccade generation or object tracking.

Conclusion

Hopefully this short overview illustrates that determining the neural mechanisms "under the hood" of direction percept in MT is still an open research question with some promising options. I believe that in order for the "correct" algorithm to be determined from direction percept depends on more than just input from areas such as V1 or V2. One must take into account a more full connectivity schema in MT where feedback from higher, downstream cortical areas may be playing a role in determining direction percept signals. Van Wezel and Britten's simple model seems like a good basis for beginning to approach this broader view of algorithmic choice in MT.

References

Born, R. and Bradley, D. (2005). Structure and function of visual area MT. *Annual Review of Neuroscience*, 28(1):157-189.

Deangelis, G. C. and Uka, T. (2003). Coding of horizontal disparity and velocity by MT neurons in the alert macaque. *J Neurophysiol*, 89(2):1094-1111.

Groh, J. M., Born, R. T., and Newsome, W. T. (1997). How is a sensory map read out? effects of microstimulation in visual area MT on saccades and smooth pursuit eye movements. *J. Neurosci.*, 17(11):4312-4330.

Grossberg, S. (1973). Contour enhancement, short-term memory, and constancies in reverberating neural networks. *Studies in Applied Mathematics*, 52, 213-257.

Nichols, J. M. and Newsome, W. T. (2002). Middle temporal visual area microstimulation influences veridical judgments of motion direction. *J. Neurosci.*, 22(21):9530-9540.

Uka, T. and Deangelis, G. C. (2004). Contribution of area MT to stereoscopic depth perception: Choice-related response modulations reflect task strategy. *Neuron*, 42(2):297-310.

van Wezel, R. J. A. and Britten, K. H. (2002). Multiple uses of visual motion. the case for stability in sensory cortex. *Neuroscience*, 111(4):739-759.

Weiss, Y., Simoncelli, E. P., and Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nat Neurosci*, 5(6):598-604.