Biological Models of Motion Perception:
Spatio-Temporal Energy Models and Electrophysiology

Christof Koch

January 26, 2006

As I mentioned in the last lecture, the Reichardt-Hassenstein correlation model of opto-motor behavior in insects has been extended to human motion perception. We’ll talk about these models and their extension, in today’s lecture.

1 Spatio-Temporal Energy Model

In the mid-eighties, a number of quantitative models of human motion perception were proposed, all mathematically equivalent to the correlation model, in that the output of neighboring receptors is linearly filtered before being multiplied (or added/subtracted and squared).

Van Santen and Sperling (1984) in their Elaborate Reichardt model introduce spatial receptive field profiles to prevent the spatial aliasing seen in the fly. This makes the response of the model zero for $\lambda < 2\Delta x$. Watson and Ahumada (1985) and Adelson and Bergen (1985) both propose models based on careful psychophysical measurements. The output can formally be expressed as a second-order interaction using the simplest type of nonlinearity possible, multiplication. Using psychophysics—which only has access to the final output of a system—none of these three models can be distinguished from each other. However, some of the internal stages different; in particular, the spatio-temporal energy model has a stage responding only to motion in one direction, while the correlation model has no such stage.

The most popular of these models, the spatio-temporal motion energy model of Adelson and Bergen (1985), carries the idea of finding the orientation of a stimulus in the x-y plane over into the spatio-temporal domain, i.e. into the x-t (or the x-y-t space). They derive a series of spatio-temporal filters that detect a certain orientation in the x-t plane. The key observation is the following.

The motion of a particular point on a moving pattern $I(x, t) = I(x-vt)$, can be represented by a straight line of slope $1/v$ in a space-time diagram. The Fourier transform (in space and time) of $I$ is then given by $I(\omega_x)\delta(\omega_x v + \omega_t)$\footnote{Remember that the Fourier transform of a shifted function, $I(x + x_0)$, is $e^{-i\omega x_0} \tilde{I}$ and the transform of $e^{-i\omega x}$ is $\delta(x + x_0)$.} In other words, the entire support of this function lies on a line of slope $v$. The Fourier transform of $I$ moving in the opposite direction is only different from zero along a line going through the origin and with slope $-v$.

One can now design filters that only have non-zero support along these diagonals to
estimate both the direction as well as the magnitude of speed. This idea can be easily generalized to 2-D motion.

Because of mathematical convenience and because the receptive fields of simple cells can be rather well approximated by Gabor functions, spatio-temporal energy models use the following pair of spatio-temporal Gabor filters with even and odd phases tuned to leftward (-) and rightward (+) motion (other oriented filters could also be used)

\[
S_{\text{even}}^\pm(x, t) = \frac{1}{2\pi \sigma_x \sigma_t} \times \exp\left(-\frac{x^2}{2\sigma_x^2} - \frac{t^2}{2\sigma_t^2}\right) \times \cos(\omega_x x \mp \omega_t t) ,
\]

(1)

and

\[
S_{\text{odd}}^\pm(x, t) = \frac{1}{2\pi \sigma_x \sigma_t} \times \exp\left(-\frac{x^2}{2\sigma_x^2} - \frac{t^2}{2\sigma_t^2}\right) \times \sin(\omega_x x \mp \omega_t t) .
\]

(2)

This generalizes in a trivial manner to 2-D space. These filters are parametrized by a particular spatial and temporal scales, \(\sigma_x\) and \(\sigma_t\).

These filters are not **separable in space-time**\(^2\). Note that the output of the correlation model is separable in space-time.

Because the output of the motion signal should not vary with time for a constantly moving stimulus and should not depend on the exact position of the stimulus with respect to the filter, Adelson and Bergen use a standard engineering trick, that of squaring and adding two filters that are 90° out of phase (based on the equation: \(\sin^2 \alpha + \cos^2 \alpha = 1\)), to render the output independent of the phase. A pair of such filters are said to be in **quadrature**. This quadratic nonlinearity also removes the dependency of the output on the sign of the contrast of the stimulus (responding equally well to a black-to-white step as to a white-to-black step edge). This amounts to computing

\[
E^+(x, t) = (S_{\text{even}}^+ \ast I(x, t))^2 + (S_{\text{odd}}^+ \ast I(x, t))^2 ,
\]

(3)

for the unit responding to rightward motion and

\[
E^-(x, t) = (S_{\text{even}}^- \ast I(x, t))^2 + (S_{\text{odd}}^- \ast I(x, t))^2 ,
\]

(4)

for the unit for leftward motion. Such an uni-directional stage is a prediction of the spatio-temporal energy models and does **not** exist in the Correlation model.

These units respond uniquely to either left- or right-ward motion. In the last stage of the model the two uni-directional outputs are subtracted from each other,

\[
E(x, t) = E^+(x, t) - E^-(x, t) .
\]

(5)

A number of psychophysical observations argue for an opponency stage. (1) It is not generally possible to see both leftward and rightward motion at the same place and time within the same frequency band. (2) Adaptation phenomena such as the motion aftereffect suggest that motion perception involves the balance between opposing leftward and rightward-movement.

---

\(^2\)A function \(f(x, t)\) is separable if it can be written as \(f(x, t) = g(x) \times h(t)\). A function expressed as a sum of separable functions, i.e. \(\sum g_i(x) \times h_i(t)\) is itself not separable.
signals. (3) It has been found that leftward and rightward-moving gratings can effectively cancel each other’s detectability.

The global output of these models, meant to mimic the motion perception of humans, is identical to the output of the correlation model formulated for the fly. That’s pretty neat!

1.1 Psychophysical Support

van Santen and Sperling (1984) present psychophysical evidence arguing that the motion system responds proportional to the product of neighbouring contrasts.

Lu and Sperling (2001) exploit two specific properties that hold for both the correlation-model in the fly and the spatio-temporal energy model of human: pseudo-linearity, that is, when a stimulus is composed of sinusoids of different frequencies, the detector response to the sum is the sum of its individual responses, and static displays are ignored, that is, the output to a stationary pattern is zero. In general, adding a stationary background pattern—the pedestal—to a moving pattern does not change the output of the motion (pedestal immunity of motion energy detectors). It has been tested in people (Lu and Sperling, 2001) and is found to be true. That it, although these pattern may look quite distinct, the accuracy of left-right discrimination for a moving stimulus and the same moving stimulus plus a stationary offset are the same.

A true speed detector, such as the gradient model in its abstract form, will either signal directly speed (analog coding) or will fire maximally at its “optimal” speed (place coding), independent of the structure of the object. In contrast, a motion detector of the correlation or energy model type does not correctly signal the local motion in terms of its velocity. Instead, its output depends on both the velocity \(v\) as well as on its spatial frequency \(\lambda\).

Diener, Wirt, Dichgans and Brandt (1976) provide psychophysical support that human observers—asked to judge the speed of a grating they fixate (when they do not track)—perceive a quantity proportional to velocity over spatial wavelength, given by

\[
M = 0.61 \frac{v}{\lambda} + v \cdot b \tag{6}
\]

The \(1/\lambda\) effect on perceived speed disappears during pursuit movements of the eye (in other words, the system that tracks smooth motion implements a different motion algorithm than the one we use to estimate motion in the scene).

Notice that much of the psychophysics carried out in support of the energy model always uses very small visual contrast values, i.e. below 5%. Higher contrast values will saturate various components of the visual pathway (e.g. we know from functional MRI in humans, that the BOLD response in area MT/MST, one of the cortical areas specialized for motion analysis, saturates above 2-3%) and thereby bring saturating types of nonlinearities into play. These higher-order nonlinearities contaminate the essential quadratic nonlinearity needed for this computation. Such an effect has been observed for the fly, where the opto-motor response

---

\[3\] The system is pseudo-linear since this property only holds for sinusoidal of different temporal frequencies.
is proportional to the square of the contrast for small values, but flattens out for higher values of contrast, as well as for the human motion system.

2 First- and Second-Order Motion

Lu and Sperling (2001) call this class of motion detectors first-order, since the raw intensities (or a linear filtered version of them) is the input. As discussed in the first motion lecture, the output of these is related to the actual optical flow of a rigidly translating object.

However, many patterns have been discovered that will to activate first-order motion detectors. An example—shown in the quicktime movie from George Mather’s website (http://www.biols.susx.ac.uk/home/GeorgeMather/Motion)—is a random dot pattern whose contrast is modulated in time, say by modulating all pixel in one or two adjacent columns reverse contrast at the same time. This does not generate a coherent motion output since the output among neighbouring pixels is not related.

This gives rise to second-order motion. It can be extracted by some spatial filtering and temporal filtering (differentiation) followed by full-wave rectification that treats positive and negative outputs the same. This function is then feed into a standard motion-energy detector.

Using clever psychophysics, Sperling and his collaborators could show that both first- and second-order motion detectors are independent and use monocular input (Lu and Sperling, 2001).

In recent years, Sperling has extended this rigorous framework to a third-class of motion detectors in humans, third-order motion. This is attentional-based and can track very complex features from one frame to the next. This is elegant since he can show that stimuli that are blind to first- and second-order motion detector can be picked up by a third-order motion system that tracks salient features in an image.

His framework is more exact and rigorous than the short- and long-range dichotomy introduced by Braddick. Short-range motion stimuli could be picked up by either first- or second-order motion detectors while long-range motion would be based on selective attention.

2.1 The Computation of Speed

It should be emphasized that both the correlation as well as the spatio-energy models do not explicitly compute motion, but rather some function which varies with the direction and the magnitude of the velocity.

Humans are pretty good when it comes to detecting small speed differences between adjacent patches; the Weber fraction for speed, $\Delta v/v \approx 0.04$ to 0.08, independent of the stimulus pattern used (for patterns above 5% contrast and in the 4 to 64 deg/s speed range; McKee and Watamaniuk, 1994).

How is the magnitude of velocity, that is speed, extracted at the neuronal level? Although
this problem still awaits a definite answer, the most likely solution exploits the fact that the Fourier transform of the function \( I(x, t) = I(x - vt) \) only has support in the \( \omega_x - \omega_t \) space along a line of slope \( v \) going through the origin.

A neuron with a receptive-field that has a support (in the Fourier) space along a diagonal of slope \( v \) going through the origin, would respond selectively to patterns of velocity \( v \), but less strongly to patterns of velocity \( v/2 \) or \( 2v \). Some sort of population coding is then used to make this computation robust and accurate (for more details see Grzywacz and Yuille, 1990).

### 2.2 Smooth versus Sampled or Discontinuous Motion

Let me briefly digress on the difference between true, continuous motion and discrete or sampled motion (i.e., the one you see on a TV or computer monitor or when watching a movie). Let me represent a pattern \( I(x, t) \) moving at a constant velocity \( v \), i.e. \( I(x, t) = I(x - vt) \). Motion of a particular point on the moving object can be represented by a straight line of slope \( 1/v \) in a space-time diagram. The Fourier transform (in space and time) of \( I \) is given by \( \tilde{I}(\omega_x)\delta(\omega_xv + \omega_t) \).

In the figure shown in class, you will see the difference between sampling a continuous moved function in \( x - t \) space versus a film of this movement: each frame remains on until the next one appears. If the sampling frequency is rapid enough, both kinds of movement will look identical (flicker fusion frequency, which depends on contrasts and color).

The difference between the sampled and the true motion represents the sampling artifact (this artifact can be negative or positive). The difference can be perceived if it contains spatio-temporal energy in the spatio-temporal range we are sensitive too (the so-called window-of-visibility).

Standard movies projectors use 35 mm film exposed at 24 frames/sec (in the early days of movie making, it used to be 16 frames/sec; thus, these films are more jerky). This gives rise to apparent motion.

Because of the unacceptable flicker rate, projectors actually use two so-called flicker blades that are added to the projector shutter. This raises the flicker frequency to an acceptable 72 frames/sec. In addition, while the film is being moved from one frame to the next, the light is cut off by a capping blade, to avoid the blurring otherwise seen. It is ironic that in order to make the action on screen appear smooth and continuous, the film must be sub-sampled (for more, see Corbin, 1969).

### 3 The Motion Pathway

As discussed early on in class, the magnocellular pathway, originating with the parasol cells in the retina and projecting into the two magnocellular layers of the LGN, is sensitive to low spatial contrasts and to high temporal frequencies.
The geniculate magnocellular cells terminate in layer 4C of V1. The output of this layer passes to layer 4B and from there directly to the middle temporal area MT. There is also a second, indirect route to MT via parts of area V2 termed thick stripes.

About 25% of cells in V1 are direction selective, that is respond strongly to motion in one, the preferred, direction and very little or not at all to motion in the opposite or null direction. V1 is the first stage where such direction selective cells—colocalized primarily to layers 4C and 4B—are found in any significant numbers.

Next to V1, no cortical area is as popular among electrophysiologists as the fifth visual area (V5 characterized by John Allman in the New World monkey and by Semir Zeki in the Old World monkey (Allman and Kass, 1971; Zeki, 1974). For all of its popularity, MT is a comparatively small part of cortical real estate, about 50 mm\(^2\) in area in the macaque, that is, about 1/20-th the extent of V1. Area MT has been identified in all primates, including humans, on the basis of four criteria: it receives a direct input from striate cortex, it contains a large proportion of cells that respond strongly to stimuli moving in one direction but show little or even a suppressed response to motion in the opposite direction, it contains a complete representation of the contralateral visual hemifield and is it is heavily myelinated.

The V1 to MT connection originates with neurons in layer 4B and, to a lesser extent, in the upper part of layer 6. This forward projection is therefore dominated by magnocellular projections. Like most connections in cortex, this one is reciprocal, with the feedback pathway from MT into V1 showing less specificity. MT receives additional input from the thick, cytochrome oxidase rich stripes of V2, containing many direction selective cells.

While one could conclude from this that all of MT input derives, directly or indirectly, from V1, this is not the case. Inactivating V1 by cooling—an operation that is reversible—or by surgically ablation, causes MT cells to respond more sluggishly and at a lower rate, yet without losing their receptive field properties (such as direction selectivity). All MT responses can be eliminated if both V1 and the superior colliculus are removed. This supports the notion that MT, as do other areas, receives a secondary visual input that bypasses the LGN by going from the retina to the superior colliculus and from there to a part of the thalamus known as the pulvinar that projects directly into MT and other visual cortical areas.

More than 80% of MT neurons prefer stimuli moving in a particular direction, with the average cell firing more than ten times stronger to motion in its preferred direction than to motion in the opposite direction. And cells retain this selectivity over a considerable range of speeds, stimulus size and position.

The moving stimulus can be a bar, a grating, or a cloud of dots. Neurons care little whether the animal moves its eye over a stationary stimulus or whether the stimulus moves. Neurons in MT respond rapidly to an appropriate stimulus, typically within 50-60 msec, compatible with the heavy myelination of this area that ensures that action potentials travel at high speed along the axons.

\(^4\)For historical reasons, MT was the label used for the area in the new world monkey, while V5 was reserved for the old world monkey. However, today both terms are used interchangeable in much of the literature, with the human homologue sometimes referred to as V5/MT.
The direction selectivity of MT is organized in a columnar fashion (like orientation-selectivity or ocular dominance in V1). Nearby neurons tend to be selective for motion in the same direction (Merigan and Maunsell, 1993).

MT provides a major output to the dorsal stream, in particular to areas MST (medial superior temporal) and VIP (ventral intraparietal).

Inactivating MT cells temporarily by injecting a chemical substance into this part of the brain increases the threshold for detecting weak motion signals in behaving monkeys without increasing the acuity, color or stereo thresholds (Newsome and Pare, 1988). This provides prima facie evidence for the critical role of MT in primate motion perception.

Historically, the cortical area in humans taken to be the homologue of monkey MT has been identified using stains that emphasize axonal myelination in cortical structures. More recently, an antibody that selectively binds to a protein associated with neurons in the magnocellular pathway have been used. Such anatomical techniques have a major drawback in that they can only be applied to post mortem material.

The advent of brain imaging has changed this situation dramatically. Positron emission tomography was first used to locate MT by Zeki, Frackowiak and their colleagues in London on the basis of its differential activation by moving clouds of dots compared to stationary ones. PET is quite sensitive to changes in blood flow dynamics in response to neuronal activity but requires injection of a radioactive bolus, hindering its repetitive use in humans. Functional MRI has no such limitation and today MT is quickly and routinely localized on the basis of the different hemodynamic responses between low contrast concentric rings that are slowly expanding and a stationary version of this stimulus or using clouds of moving stimuli. We’ll demo this in class using a little movie from David Heeger’s lab.

Now let me discuss single cell evidence that supports the hypothesis that neurons in V1 and MT underlying the perception of motion.

4 Spatio-Temporal Oriented Receptive Fields in V1

There exists good electrophysiological evidence in from cat primary visual cortex, that directional selective cells there behave as if they compute uni-directional motion energies (McLean and Palmer, 1989; Reid, Soodak and Shapley, 1991; DeAngelis, Ohzawa and Freeman, 1993).

Mclean and Palmer looked at 52 simple cells in cat area 17. They found that 16 out of 24 of the cells they tested that were spatio-temporal separable did not respond preferentially to motion in one or the other direction. 8 out of these 24 cells did show some direction asymmetry. All of the remaining cells that they encountered (28/28) had oriented subregions. And these showed directionally asymmetric responses to moving stimuli. The optimal speed of a moving bar could be predicted from the slope of the orientation in the x-t receptive fields.

In a detailed comparison of the correlation and the energy models with extracellular recorded responses of five complex cells from cat striate cortex, Emerson, Bergen and Adelson (1992) provide direct evidence that these complex cells (i) show spatio-temporal orientation
and (ii) implement an uni-directional stage. Although this cell count is very low, this does provide solid evidence in favor of the motion energy model. Interestingly, none of their cells implemented the opponency stage.

The experiments carried out by Snowden, Treue, Erickson and Anderson (1991) point towards the existence of neurons in V1 that signal the direction of motion in rough qualitative agreement with the unidirectional motion stage in the spatio-temporal energy models. They also provide evidence that the opponency suppression stage could be carried out within MT neurons.

When recording from simple cells in the cat visual cortex, Movshon and Tolhurst (1975) systematically varied the temporal frequency \(\alpha = v/\lambda\) (how many periods of the grating move past a particular point per second) of a drifting sinusoidal grating and find that the tuning curve has a single peak at an optimal value of \(\alpha_{\text{optimum}}\). They then repeated this experiment using a different spatial wavelength \(\lambda\) and found that the peak of the response was again at \(\alpha_{\text{optimum}}\). This is to be expected, since a neuron implementing the spatio-temporal energy model should be sensitive only to \(\alpha\) (put differently, plotting the neuron’s response as a function of speed \(v\), different \(\lambda\)'s should yield a peak of \(f(v)\) at different values of \(v\)).

Newsome and Movshon (1981) reported in an abstract that some fraction of neurons in area MT of the monkey respond differently: similar to V1 cells, their firing rate \(f\) as a function of temporal frequency \(\alpha\) has a single peak. However, the peak of \(f(\alpha)\) shifts with different values of \(\lambda\), in agreement with the hypothesis that these neurons are sensitive to velocity \(v = \alpha \lambda\) (or, put differently, their firing rate as a function of velocity \(f(v)\) should have a single peak, independently of the spatial wavelength \(\lambda\) used).

5 Motion Integration in MT

An elegant psychophysical and electrophysiological experiment supports a two-stage model of motion computation: a local stage for motion registration and a more global one for motion integration.

Adelson and Movshon (1982) superimposed two moving square gratings of different orientations to produce the percept of a “pattern” moving in a single direction, with the direction being determined by the direction of the two components. Both gratings move at the same speed at right angles to each other. Under this condition, humans do not observe the individual gratings moving but perceive the resulting pattern as a single coherent plaid. The direction of the resultant plaid pattern (termed pattern velocity) can be computed from knowledge of the local component velocities of the two gratings; in fact, the resultant coherent motion satisfies the smoothness constraint operating in the Horn and Schunck formulation of the optical flow computation, such that the final optical flow is compatible with the two component velocities and is constant over the image. As long as the two gratings are rather similar to each other, we perceive coherent motion. However, if the contrast between the two gratings differs by an order of magnitude or more or if the spatial wavelengths differ by an order of magnitude or more, observers perceive two two gratings sliding past or over each other (Adelson and Movshon, 1982).
These plaid experiments were repeated with monkeys as subjects while electrodes recorded the response of cells in V1 and MT (Movshon, Adelson, Gizzi and Newsome (1985). All neurons in V1 and about 2/3 of all cells in MT only responded to the velocity component associated with either one of the two gratings. These are the component selective neurons. However, some fraction of cells in MT—termed pattern-selective neurons—responded best to the motion of the coherently moving plaid pattern, and not the motion of the two individual gratings. The prevalent interpretation of this experiment is that this fraction of cells in the deep layers of area MT integrate motion information, thereby solving the aperture problem. Note that the response of these MT neurons—rather than the V1 cells—is in agreement with our perception. This is one instance of the general rule that the response properties of higher cortical areas corresponds better to our perception than the responses of lower areas.

Furthermore, note that pattern-selective neurons are nonlinear with regard to the response to their individual components. The neuron does not respond to the entire pattern as it would to the individual components.

6 Motion Perception and Brain Activity

These observation provided the rationale for a series of powerful experiments that link the firing rates of individual MT cells while the animal is carrying out a motion discrimination task. The experiments were conceived and carried out over a period of several years by the neurobiologist William Newsome at Stanford University. Although his paradigm is not an easy one to follow, it is well worth our time (Newsome, Britten and Movshon, 1989; Britten, Shalden, Newsome and Movshon, 1992).

The motion stimulus consisted of a cloud of bright dots moving randomly on a monitor. What varied during the course of the experiment was the strength or coherence $c$ of the motion signal. For $c=100\%$, all the dots move in unison in one direction, called the preferred direction. As $c$ is reduced, a progressively smaller portion of dots translate coherently in the preferred direction while the rest move in a random way in any which direction. For instance, at $c=50\%$, half the dots move in their common direction while the remaining half of the dots jitter about. Humans see the coherent motion in the preferred direction overlaid onto motion noise. For $c=0\%$, no net motion component is left, since any dot can move in any direction from anywhere on the screen, creating the visual appearance of a snow storm with the wind blowing snow crystal everywhere.

Newsome and colleagues trained the animal—using a binary task procedure—to recognize motion at very low contrast levels and computed the associated psychometric curve for this task. The observer, human or monkey, has to decide for any particular motion pattern—displayed for 2 sec on the monitor—whether the coherent motion was in the preferred or in the null direction. This is an instance of what psychologists call a binary task. The subject is given no opportunity to say “I don’t know,” or, “I didn’t see the stimulus,” but has to respond in one of two ways: “the coherent dot cloud moved up” or “to down”. From the methodological point of view, this make interpreting the results much easier.

Obviously, this is terribly easy to do if all dots move coherently in one direction. But if
the motion signal is impaired by reducing its strength, one has to pay considerable attention and begins to make mistakes. At \( c = 0 \), the observer has to guess and, if the experiment is repeated over and over and is unbiased, will only get the right answer—by chance—on half of all the trials. Threshold performance is defined to be the smallest amount of coherence that permits the observer to be right 82% of the time. The psychometric curve is simply a record of how well the observer did at this discrimination task as the motion coherence is varied.

In order to link neuronal firing behavior to perception, Newsome and colleagues postulate that performance depends on a comparison between neurons selective to motion in the two opposing directions. The performance of a MT neuron can then be characterized as the probability that a randomly selected response from the preferred direction is larger than a randomly selected response from the null direction distribution. This decision rule chooses the correct direction on about half the trials at 0.8% correlation (i.e. the monkey is essentially guessing), whereas it performed nearly perfectly at 12.8%. Using ROC signal detection theory to estimate this threshold at each correlation level results in a neurometric curve which can be superimposed onto the psychometric curve. Evidently the two curves are very similar; in fact in this curve, the neuronal threshold was slightly lower (4.4%) than the psychophysical one (6.1%). Analyzing this data for more than 100 MT cells, Newsome and his colleagues conclude that the neurometric curve of half of the cells was virtually indistinguishable from the psychometric ones, while the neuronal threshold in the remaining cells was lower than the psychometric one. In other words, were the monkeys able to select and measure the discharge of some of these neurons as accurately as the experimentalists did, their performance could have been better than it actually was. While this may appear paradoxically, it can be explained by postulating uncertainty in the decision process as well as summating over many, partially correlated neurons (Britten et al., 1992). The monkey is uncertain about the nature of the stimulus to be discriminated and must base his judgment on the pooled activity of a potential large number of partially correlated neurons, instead on a single one.

What should be pointed out here is that Newsome's estimate of the neuronal threshold is based solely on the total number of spikes discharged during the 2 sec long duration of any trial.

Let me hasten to add that nobody is claiming that the behavior of the animal is caused by the discharge of a single MT neuron (out of a few million) nor that the cells postsynaptic to the recorded neuron infer the direction of motion using this particular signal detection technique. Rather, the claim is that by observing the cell’s behavior over many trials (enough to collect relevant statistics), it is possible to derive a quantitative one-to-one relationship between the number of spikes generated by the stimulus within the trial period and the behavior of the animal.

Nobody would be surprised if, upon recording from the motorneurons controlling the muscles that are used by the monkey to indicate its behavior, one finds a perfect correlation between their activity and the animal’s behavior. But, in the Newsome experiment, the cells were located deep in visual cortex and were not in any way directly linked to the output system.

This finding certainly strengthen the conviction of those electrophysiologists who argue that it does not necessarily follow from the highly distributed nature of cortical processing that
one must record of necessity from many neurons simultaneously. Listening in on individual neurons has, and will in all likelihood continue to yield, great insights into the way the brain works.

7 Influencing Perception with an Electrode or How to Go From Correlation to Causation

Based on the classical brain stimulation experiments by the neurosurgeon Penfield (Penfield and Perot, 1963), Newsome was casting about for more direct ways in which MT activity could be shown to cause motion perception. Based on the evidence for brain stimulation laid out in this chapter, he reasoned that one might be able to systematically influence the animal’s behavior in a motion discrimination task by exciting neurons in MT. The big question was whether or not microstimulation of a tiny piece of brain would show up (for the detailed story, see Salzman et al., 1990, 1992).

If neurons selective for the different directions of motion are randomly distributed in MT, then stimulating any small group of adjacent cells is unlikely to give rise to a net signal. If one simultaneously excites cells that code for motion to the right and other cells that code for leftward motion, then one might not expect any effect.

Cortical area MT is organized, however, like other brain areas in columns. Columns for both direction of motion and for depth. And if one could restrict the effects of electrical stimulation to neurons within a column, one might expect that injecting current into a column whose neurons code for rightward motion makes the animal more likely to decide that motion did occur towards the right. And this is exactly what was observed.

The same electrode that recorded the cell’s discharge was used to deliver a high-frequency train of biphasic electrical current pulses. Newsome estimates that these 10 μA current pulses directly activated neurons within a sphere roughly 140-240 μm in diameter. This volume contains somewhere between 100 and 1,000 nerve cells.

Trials on which current was delivered were randomly intermixed with those during which no stimulation occurred. The transparency exemplifies results from four stimulation experiments. The psychometric curves obtained during stimulation are clearly shifted, to a variable extent, to lower coherence values compared to the curves obtained without tickling MT. That is, when MT was stimulated, the animal was more likely to respond to motion in the cell’s preferred direction than otherwise. The effect of stimulation can be quantified by the change in threshold, which range between 5.6% and 30.6%.

Put differently, stimulating the MT cell and its neighbors is equivalent to adding between 5 and 30% dots to the motion displays. Or, behaviorally, a motion stimulus of coherence $c$ is indistinguishable from a weaker motion stimulus supplemented by electrical current.

Note that these effects were highly specific. Stimulating sites whose neurons had receptive fields that did not overlap with the receptive field of the neuron recorded from did not change the animals behavior. Presumably because a neighboring columns was activated.
Sometimes during a single session, two sites could be stimulated, sites whose direction of motion were opposite each other. And, as expected, stimulation in the two sites had opposite effect on behavior. Furthermore, stimulating MT did not influence motor behavior by itself.

What Newsome and his colleagues infer from all of this is that it is likely that the perception of the animal was affected in a very specific and topographic manner by the synthetic electrical stimulus. They further concluded that MT is directly and causally involved in the generation of motion perception.

I would dearly love to know what the monkeys saw during stimulation. Did MT stimulation manifest itself in some explicit manner by visibly changing the perception of the cloud of moving dots? That is, did the monkey observe phosphenes? Or was the effect implicit and remained unconscious; that is, the animal somehow felt it had to respond in that particular direction without clearly seeing something? We will not know the answer until such experiments are repeated on humans, possibly as part of an invasive neurosurgical procedure.

8 References


