**PHENOMENA OF MOTION PERCEPTION**

1) What is motion? Apparent motion and real motion
2) Long-range motion:
   - Arguments for short- and long-range mechanisms
3) Korte’s laws and figural affinity

---

**UTILITY OF MOTION PERCEPTION**

Nakayama (1985): Tasks for which motion processing is useful:

1) Encoding 3-D shape
2) Perceiving time-to-collision
3) Segmentation
4) Proprioception
5) Control of eye movements
6) Pattern vision
7) Perception of moving objects

How these tasks are accomplished is not well understood!

---

**UNITS OF MOTION PERCEPTION**

J. J. Gibson (1979) warns that our “first order” intuitions about the origins of spatiotemporal patterns of visual input in the environment may lead us astray.

Many assume that “the problem” of motion perception is one of successively, over time, inverting the projective correspondence between a patch of a physical surface and its projection on the retina.

However, many of our most vivid and ecologically pertinent experiences with motion -- particularly those involve with occlusion and disocclusion over time -- do not conform to such an analysis.

---

After reviewing some of the perturbations of the optic array caused by observer motion or external change, Gibson says:

Can these disturbances of structure be treated mathematically? They surely cannot all be treated with the same mathematical method, for some of them do not conform to the assumptions of the theory of sets. Some of the above changes do not preserve a one-to-one mapping of units over time, inasmuch as the [optic] array gains or loses units in time. Accretion or deletion of texture during occlusion is one such case. Foreshortening or compression of texture preserves one-to-one mapping only until it reaches its limit, after which texture is lost. The emergence of new texture with rupturing of a surface, the nullification of texture with dissipation of a surface, and the substitution of new texture for old are still other cases of the failure of one-to-one mapping, or projective correspondence. In all of these cases it is not the fact that each unit of the ambient array at one time goes into a corresponding unit of the array at a later time. The case of an optic array that undergoes “flushing” or scintillation of its units is another example, and so is what I called fluctuation in connection with changing light and shade. [Gibson, 1979, p. 108; emphasis added]
Arguments for motion as a fundamental visual modality
(not just "derived" as displacement across time):

Motion aftereffects
Single-cell recordings
Phi motion over distances irresolvable by static perception

http://psychlab.psych.purdue.edu/MagniPhi/ARVODemo.html

Explore this great web site; learn how the phi phenomenon launched the Gestalt revolution!

EXNER (1875): MOTION AS A PRIMARY MODALITY

1) Simultaneous perception of motion in two directions rules out explanation by eye movements.
2) Motion is perceived for dot displacements that are below hyperacuity thresholds; therefore is not "derived" from displacement over time.

Beta motion: continuous, "optimal" object motion
Phi motion: "pure" movement, "objectless"
Gamma motion: at onset of single flash

NOTE:
First-order is not the same as short range.
Second-order is not the same as long-range.

Let's start with long-range motion!

INTERSTIMULUS INTERVAL

ISI: Interstimulus Interval
SOA: Stimulus Onset Asynchrony

*Note that the definition of ISI in the animal learning literature differs, corresponding to what is called SOA above!
**KORTE’S “LAWS” FOR LONG-RANGE MOTION**

S: Spatial separation  
ISI: Inter-Stimulus Interval  
I: Stimulus energy (luminance)

For optimal apparent motion:
1) For fixed ISI: \( S \sim I \)
2) For fixed S: \( I \sim 1 / ISI \)
3) For fixed I: \( S \sim ISI \)

**KORTE DATA**

For fixed I: \( S \sim ISI \)

Note that both high and low cut-offs for apparent motion are shown.

The proportionality above is best seen on the low cut-off for the 10 msec curve.

**FIGURAL AFFINITY 1**

Figural affinity in split apparent motion (Ullman, 1979)

First frame:  
Second frame:  

Closer is preferred.  
Same orientation is preferred.  
Same length is preferred.

**FIGURAL AFFINITY 2**

Same luminance is preferred.

Claim: Same topology is preferred. ???  
Li, 19???
**STUART ANSTIS**

http://psy.ucsd.edu/~sanstis/motion.html

http://www.psy.vanderbilt.edu/faculty/blake/214_F2001/W14Demos/MotionDemos.html
(Randolph Blake demo page)

---

**KINDS OF LONG-RANGE APPARENT MOTION**

**Beta:** Continuous motion of a well-defined object across empty intervening space

**Phi:** Sense of motion without a concurrent perception of moving object

**Gamma:** Apparent expansion at onset or contraction at offset of a single flash of light.

**Delta:** Beta or phi motion directed toward the first flash, when the intensity of the second flash is sufficiently greater than the first

---

**TWO-FLASH APPARENT MOTION DATA**

**Key facts:**

A single flash (stationary) does not generate a motion percept (gamma motion aside).

Two or more properly timed and positioned flashes do.

A change of ISI (only) can alter the global motion percept qualitatively (e.g. Ternus effect)

**Issues:**

Why is the long-range influence of a single flash -- which must exist -- not perceived as motion?

How do individual flashes interact to bridge variable distances without a loss of sharpness in the percept?

How does varying ISI cause a speed-up or slowing down of the smooth motion signal?

---

**TRAVELING G-WAVES**

1) Contemplate the stages labeled “Level 1, Level 2, ..., Level 5” in the Grossberg & Rudd (1989) Moving Oriented Contrast (MOC) Filter.

2) Forget them.

3) Remember the “level-less” long-range Gaussian filter, whose traveling Gaussian (G) waves do all the work in long-range apparent motion.

G-waves are also implicated in the spatial deployment of visual attention!
**SINGLE FLASH SPATIAL PATTERN**

If a single narrow peak is produced at Level 4, the Gaussian filter produces a Gaussian profile, which is then sharpened back to a single peak by a “winner-take-all” competition at Level 5.

Note 1: The width of the Gaussian varies, depending on the (size of) the “scale” considered.

Note 2: A Gaussian of any scale changes amplitude over time.

**TEMPORAL PROFILE OF SINGLE FLASH**

A single flash creates a characteristic rapid increase toward saturation level while ON and a corresponding exponential decay soon after it shuts OFF.

This temporal profile modulates the amplitude of the Gaussian signal.

Since the peak of the Gaussian stays in place regardless of amplitude, the model’s “percept” is that nothing moves.

**TEMPORAL PROFILE OF TWO FLASHES**

If two flashes occur in rapid succession, the “waning” of the first signal may overlap in time with the “waxing” of the second.

Note: To produce apparent motion, it is not necessary that the temporal profiles of the stimulus “abut” (i.e. zero ISI).

ISI could be positive, or the flashes themselves could overlap in time somewhat.

**TRAVELING WAVE: LONG-RANGE MOTION**

If the Gaussian activity profiles of two flashes overlap sufficiently in space and time, the sum of Gaussians produced by the waning of the first-flash Gaussian, combined with the waxing of the second-flash Gaussian, can produce a single-peaked traveling wave.

The wave is then processed through a “choice” network.

The resulting motion percept is thus both long-range and sharp.
**“JUST IN TIME” DELIVERY OF MOTION SIGNALS**

For a given ISI:

How does perceived velocity increase with distance between flashes?

For Gaussian filter:

\[ G_{ji} = H \exp \left\{ -\frac{(j-i)^2}{2K} \right\} \]

the largest separation, \( L_{crit} \), for which sufficient spatial overlap between two Gaussians centered at locations \( i \) and \( j \) will exist to support a traveling wave of summed peak activity is:

\[ L_{crit} = 2K \]

---

**NO MOTION VS. MOTION**

Interactions between flash spatial separation and filter scale

*Multiple spatial scales!*

To upper left of dashed lines: **NO MOTION**

Why no motion?

The waxing and waning Gaussian profiles never produce a single peak of summed activation.

To lower right of dashed lines: **MOTION**

---

**EQUAL HALF-TIME PROPERTY**

Not only does the traveling wave for a given Gaussian scale bridge variable distances in (almost) equal times, but also:

Traveling waves from Gaussian filters of a variety of sizes bridge the same distance in comparable times; in fact, the time needed to bridge half the distance between flashes is precisely the same:

Therefore, no “conflict” in motion percept across multiple scales!

---

**THE UNITS OF MOTION PERCEPTION**

Issue: Are we first finding features, and then matching or tracking them over time?

OR

Are we extracting “motion energy” from the spatiotemporal optical “ooze,” before we conclude that anything is moving?
SHORT-RANGE MOTION DEMO

http://www.bu.edu/smec/lite/perception/camouflage/color.html

Note:

Absence of static information for form.

Accretion/deletion at edges of moving form.

Sharpness of perceptual edges.

RANDOM DOT KINEMATOGRAMS: D-MAX

“Short-range” apparent motion occurs for stimuli composed of dense arrays of small elements (e.g. dots).

Question: What is the “D-max” for “short-range” apparent motion? I.e., what is the largest tolerated displacement for corresponding dots in a dense kinematogram for seeing motion (by “sensing” correlation across two frames)?

MOTION LONG AND SHORT

Table 1 from Cavanagh & Mather (1989) -- See Syllabus (principally after Anstis (1980) and Braddick (1980), with some additions).

<table>
<thead>
<tr>
<th>Short-range</th>
<th>Long-range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short spatial range (&lt; 15 arc min)</td>
<td>Operates over many degrees</td>
</tr>
<tr>
<td>Braddick (1974) [<em>D-max</em>]</td>
<td>Kolers (1972)</td>
</tr>
<tr>
<td>Brief temporal range (80-100 ms ISI)</td>
<td>ISI up to 500 ms</td>
</tr>
<tr>
<td>Motion aftereffect</td>
<td>No motion aftereffect</td>
</tr>
<tr>
<td>Not dichoptic</td>
<td>Dichoptic</td>
</tr>
<tr>
<td>Braddick, 1974</td>
<td>Shipley et al. (1945)</td>
</tr>
<tr>
<td>No response to colour</td>
<td>Response to colour</td>
</tr>
<tr>
<td>Low-level neural comparator</td>
<td>Responsive to higher-level correspondences that do not activate motion detectors</td>
</tr>
<tr>
<td>Passive motion response, velocity space computations</td>
<td>Cooperative processes, inference</td>
</tr>
</tbody>
</table>

Cavanagh & Mather, 1989 is an important review paper that helped to define an emerging zeitgeist.

The phrase “velocity space” in the previous panel refers to a method for combining estimates of motion directions for displays containing two or more component motions. We will treat this topic next week. For now, note simply that this phrase connotes one approach among several to the problem of motion direction estimation.
MAD (D-)MAX: THE REVENGE OF MULTIPLE SPATIAL SCALES

Later studies showed that D-max can be much more than 15 arc min if:

1) More than two frames are used, or
2) Low spatial frequencies are used

D-max can be thought of as (more like) 1/4 of the spatial “duty cycle” of a “balanced” periodic stimulus:

e.g.

```
+---+---+---+---+---+
|   |   |   |   |   |
+---+---+---+---+---+
```

frame 1

```
+---+---+---+---+---+
|   |   |   |   |   |
+---+---+---+---+---+
```

frame 2

vs.

```
+---+---+---+---+---+
|   |   |   |   |   |
+---+---+---+---+---+
```

frame 1

```
+---+---+---+---+---+
|   |   |   |   |   |
+---+---+---+---+---+
```

frame 2

FIRST-ORDER AND SECOND-ORDER STIMULI

First-order and second-order motion stimuli are sometimes referred to as Fourier and non-Fourier stimuli, respectively.

A second-order motion stimulus is one in which the “thing” that moves has the same mean luminance as a stationary background -- over a sufficiently large spatio-temporal averaging window.

The local contrasts that make up the moving “thing” may themselves scintillate (change sign of contrast) so that no local element “persists” long enough to be tracked.

http://www.cnl.salk.edu/~maarten/demos/2nd.html

FIRST-ORDER AND SECOND-ORDER PROCESSES

Cavanagh and Mather, 1989

First-order stimuli engage earliest levels of processing (e.g., retinal ganglion cells); contain spatiotemporal correlations in luminance color (wavelength).

Second-order stimuli are defined by frequency with which specific combinations of intensity or color values occur for pairs of points.

Same mean luminance and color (wavelength) form a spatiotemporal “texture” or, pairs in different time frames of motion sequence create binocular disparity (depth).

SECOND ORDER AGAIN

There are many kinds of second order stimuli a.k.a. “non-Fourier” stimuli

What they have in common is that no first-order detector, of sufficient “size” (in space for texture, space-time for motion) would respond to them . . . true if the mean luminance of the “black” and “white” squares on the right is the same as the luminance of the background.
**SECOND SECOND-ORDER EXAMPLE**

Another kind of second order motion:

- time interval 1
- time interval 2

Perceived motion of “region of downward moving dots” is to the right.

(The dashed boxes are not part of actual display; dots disappear when they collide with virtual boundary.)

---

**D-MAX REVISITED**

Considering first-order and second-order stimuli, spatial scales, multiple frames vs. two frames, etc.

**D-max** seems to be less of a fixed (retinal) quantity and more dependent on stimulus characteristics (e.g., contrast polarity, spatial frequency, density of elements in a display).

Consider analogies to static grouping “D-max”:

---

**SHORT AND LONG; FIRST AND SECOND**

Cavanagh & Mather argue that long-range vs. short-range is less important than first-order vs. second-order, and that pools of detectors of either type could exist for different spatial scales.

Note that the diagram on the right includes (spatial) gradient operators* as front ends of Reichardt-type* “delay and compare” units!

---

* Defined later in this lecture. The point is that these diagrams achieve “politically neutrality” by blending elements of rival models.

---

With thanks to Lavanya Viswanathan for development of some of the material on subsequent pages.
**GENERIC MODEL OF MOTION DETECTION**

"Skeletal model" (Nakayama, 1985)

early stage: sensitive to position and spatial frequency

second stage: directionally selective subunits (e.g., Riechardt detectors)

third stage: spatial and temporal integrator

---

**FRONT END**

Why not just use the change of *local image intensity* over time to encode the motion of the surface?

By doing so, for ramp luminances, *motion of the source of light* would yield equivalent results to *motion of the surface* itself.

The motion system would be contaminated by extraneous changes in ambient illumination.

---

**ISSUES UNRESOLVED BY SKELETON MODEL**

Mechanism for *directional selectivity* is unspecified: addition/thresholding? multiplication? inhibition ("nulling" or "veto" mechanism)?

Contrast polarity: Must *sign of contrast* at edges *match* over time in order to stimulate detectors?

What is the output of the system? *Direction* of motion? *Metrical speed*? *Normal component of motion* (ref., the *aperture problem* -- to be discussed next week)?

---

**CORRELATIONAL MODELS: REICHARDT DETECTOR**

A *Reichardt detector* (Reichardt, 1961) consists of a pair of receptors separated by some physical *distance* such that the *delayed output* of one receptor is *multiplied* by the output of the other receptor.

This detector is the acknowledged early *motion mechanism* for fly vision.
CORRELATIONAL MODELS: NULLING VIA INHIBITION

Barlow and Levick, 1965:

Two receptors connected to an AND-NOT gate, one directly and the other through a delay.

Active inhibition of the non-preferred direction occurs.

A spot arriving at R1 before R2 will cause the inhibition due to R1 and the excitation due to R2 to arrive at the gate simultaneously, thus shutting off (nulling) the output.

Such a mechanism is found in rabbit retina.

CORTICAL NULLING?

Barlow and Levick originally proposed the nulling model to explain the directionally selective responses of cells of the rabbit’s retina.


Meynert cells have asymmetrical dendrites that are believed to make them sensitive to motion in one direction.

MEYNERT CELL DIRECTIONAL SELECTIVITY

Inputs:
Basal dendrites: excitation from a different part of the visual field
Cell body: inhibition from the current position in the visual field

Velocity sensitivity:
preferred velocity = \( \frac{\text{spatial displacement of two parts of receptive field}}{\text{relative delays of excitatory and inhibitory responses}} \)

CAJAL PLUS ONE CENTURY

Livingstone, 1999, Neuron, 20,509-526; Fig 9
DIRECTIONAL SELECTIVITY THROUGH NULLING

Figure 2. The possible organization of a direction-selective cell. (From ref. 5.) [*] Flashing a bar in the left side of the receptive field (upper left) evokes an excitatory postsynaptic potential (EPSP), whereas flashing the same bar in the right side of the receptive field (lower left) evokes an inhibitory postsynaptic potential (IPSP), but with a longer delay. When a moving bar encounters the receptive field from the left (upper right), it first triggers the EPSP from the left side of the receptive field, and only later, the IPSP from the right side. The EPSP and IPSP sum at the cell body (E+I), but the IPSP is too late to prevent the EPSP from evoking action potentials in the cell. The bar approaching from the right, however, has a very different effect (lower right). If the bar’s speed is in the right range, it will encounter the right side of the receptive field about 25 ms before it reaches the left side, giving the IPSP time to reach the cell in synchrony with the EPSP. The resulting sum of the EPSP and IPSP (E+I) is much smaller than for leftward motion, so that the IPSP effectively prevents the EPSP from generating action potentials in the cell.

ELABORATED REICHARDT DETECTOR (ERD)

Another type of correlational model is the Elaborated Reichardt Detector (ERD) van Santen & Sperling, 1985:

A modified Reichardt detector with band-limited spatial frequency channels at the front end . . . eliminates the spatial aliasing of the original Reichardt model and corrects the wrong prediction of a reversed motion percept to any continuously moving object that is dominated by spatial frequencies whose half period is smaller than the inter-detector spacing.

SF: spatial frequency filter
TF: temporal frequency filter
i.e., delay
TA: temporal averaging

ISSUES WITH CORRELATIONAL MODELS

<table>
<thead>
<tr>
<th>Good news</th>
<th>Bad news</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple!</td>
<td>Too simple?!</td>
</tr>
<tr>
<td>Provide speed sensitivity (in some ranges)</td>
<td>Identical outputs for fast motion in the preferred direction and for slow or stop-start motion in the non-preferred direction</td>
</tr>
<tr>
<td>But the human visual system is also subject to the correspondence problem of apparent motion!</td>
<td>Need to match features.objects at two spatial locations, i.e., correspondence problem of apparent motion</td>
</tr>
<tr>
<td>Cannot explain contrast-polarity-specific motion</td>
<td>Cannot explain second-order motion (without “preprocessing”)</td>
</tr>
</tbody>
</table>

CORRESPONDENCE PROBLEM

Under stroboscopic illumination, a fast clockwise turning wagon wheel can seem to be turning slowly counter-clockwise.

Implicit Assumption: the nearest spoke in the next frame is the same one as in the last frame.
**GRADIENT MODELS**

Marr and Ullman, 1981: Convolv image with the Laplacian of a Gaussian, $\nabla^2 G \ast I$, to find the edges in the image (zero-crossings of the Laplacian), and then use spatial frequency filtering to detect motion.

Mathematical veridical velocity, under constant illumination, is given by:

$$V_x = \left( \frac{dl}{dt} \right) / \left( \frac{dl}{dx} \right)$$

Marr-Ullman motion estimator:

$$V_x = \frac{\partial}{\partial t} \nabla^2 G \ast I$$

In other words, compare the sign of spatial contrast at an edge (“bright to dark,” or “dark to bright”) to whether luminance is increasing or decreasing near that edge.

---

**MARR/ULLMAN MOTION DETECTOR**

Claim*:

- **X cells** signal $\nabla^2 G \ast I$, the presence of a zero-crossing and its direction of contrast.

- **Y cells** signal $\frac{\partial}{\partial t} \nabla^2 G \ast I$, the time derivative of luminance near the zero-crossing.

However, Y cells get rectified inputs and are unlikely to transmit the sign of the derivative.

Might something else do this job?

*Recall: “X” and “Y” are cat cell types.

---

**X AND Y COMPUTATIONS**

**Increase** in luminance over time in a region where a high-to-low luminance edge exists is consistent with left-to-right motion.

**Decrease** in luminance over time in a region where a high-to-low luminance edge exists is consistent with right-to-left motion.

---

**ISSUES WITH GRADIENT MODELS**

**Good news**

- Provide *speed sensitivity*

- *No correspondence problem;* output depends only on one point in the image, i.e., -- can respond instantaneously -- sensitive to small displacements

- Explains contrast-sensitive-motion and luminance adaptation effects

**Bad news**

- Humans experience *correspondence problem* under strobed illumination

- Cannot explain *long-range effects*

- Cannot explain *second order motion*

- Biological implementation?
PHENOMENA FAVORING GRADIENT MODELS

Gradient models are compatible with data on contrast-polarity-specific adaptation effects. (Moulden & Begg, 1986.)

Gradient models: “Unilocal” sensitive to adaptation (luminance)
Correlation models: “Bilocal,” work best (only?) at low contrast

Frame 1: activates sustained “bright/dark” cell
Frame 2: activates “luminance decreasing” unit
Percept: leftward motion

<table>
<thead>
<tr>
<th>T</th>
<th>S</th>
<th>BD</th>
<th>DB</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>--</td>
<td>←</td>
<td>←</td>
</tr>
</tbody>
</table>

S: sustained
T: transient
BD: bright-dark contrast
DB: dark-bright contrast
+: increase in luminance
--: decrease in luminance

SPATIO-TEMPORAL ENERGY MODELS

Adelson and Bergen (1985) proposal:

Feature-match models: Global-match models:

1) locate features in two frames
2) establish correspondences between features
3) determine $D_x, D_t$ between frames
4) compute $D_x/D_t$

Compute a global cross-correlation between frame regions, finding the $D_x$ that gives the best match between frames (whose time difference, $D_t$, is implicitly known).

Adelson and Bergen argue for global match models.

ADELSON AND BERGEN, 1985 (CONT’D.)

Moving bar: few features to match

Moving random stripes: many features to match

REAL VS. APPARENT MOTION

Motion as “orientation in space-time”

a) Moving bar
b) Real motion
c) Sampled motion

Spatio-temporal receptive fields can respond to sampled or real motion, (equally well if sampling interval is below resolution thresholds).

Consider this as an alternate view of short-range “apparent” motion.
SPACE-TIME SEPARABILITY

Filters oriented in space-time seem to require “a different temporal impulse response correctly tailored to each spatial position within the receptive field.” (A & B, p. 288)

Such filters would not have separable responses in space and time, unlike the one on the right, that does.

MOTION AND SEPARABILITY

Space-time separable filters can respond to motion, but directional selectivity is compromised.

DIRECTIONAL SELECTIVITY

Separable filter (odd-symmetric)

Space-time oriented filter (inseparable) -- selective for rightward motion

FROM SPACE TIME SEPARABILITY TO INSEPARABILITY
PHASE SENSITIVITY

Spatio-temporal energy filters are inherently phase-sensitive. (i.e., their response fluctuates with the momentary alignment of a pattern with the spatial cross-section of the receptive field that is “operative” at that instant. In other words, their response depends on the sign of stimulus contrast at a given moment.)

A & B recommend finding a phase-insensitive response by computing motion energy: by squaring and summing outputs of “quadrature pair” filters.

ENERGY AND REICHARDT MODELS

CORRELATIONAL VS. ENERGY MODELS

In a 1984 paper on “elaborated Reichardt detectors” (ERDs), van Santen and Sperling recommend that we dispense with the classical conception motion as involving a spatial object occupying different locations at different times, and instead “... think of motion as involving a temporal object (luminance modulation pattern that occurs at different points in time at different locations) (vS & S, 1984, p. 451)

In 1985 they prove that energy models and ERDs are formally equivalent, in that they compute identical outputs for any given input!

A & B dispute this assertion, claiming that the identity only holds for a special case of energy models, and later argue that physiological evidence concerning intermediate stages favors their approach.

PARTING THOUGHTS

1) Ask Ennio about papers arguing that gradient schemes and correlational models are special cases of a single scheme, whereby comparisons among spatial neighbors of local correlational units display: “gradient-like” processing for low-noise conditions, and “correlation-like” processing for high-noise conditions.

2) Models considered to date only handle (at best) “low-level” motion detection in low-noise conditions, or in conditions where noise is “unbiased.” BUT . . . issues of segmentation and grouping, occlusion, “border ownership,” etc. pervade motion processing, just as in static form vision. These will be considered next week.