INI 402 - Computational Vision
Script

Samuel Gunz[1] and Martin Emons[1]

[1]Department of Biosystems Science and Engineering, ETH Zurich

Spring 2020

1 Visual Pathways

The peripheral sensory organ of the visual system is the eye. The visual field can be partitioned into a *monocular* (only visible with one eye) and a *binocular* (made up of both eyes) part. The retinal image is inverted compared to the visual field, meaning that the left part of the visual field is on the right side of the retina and vice versa. The left half of each retinal field continues left at the *optic chiasm*, whereas the right halves continue right. The parts that change over cross are called *controlateral*. The *lateral geniculate nucleus* (LGN) is a place where most optic axons terminate, whereas a small portion continues until the superior colliculus. The axons are then oscillating between the LGN and the *striate cortex* and enters then the *primary visual cortex*. The primary visual cortex is partitioned in a right and a left primary visual cortex, where there is a lower and upper quadrant.

Anatomy of the Eye The retina is a part of the *central nervous system* (CNS). The retinal already makes some preprocessing and discards some of the information. The fovea is the point of best vision (later in more detail) and the blind spot is where the optic nerve leaves the eye. From the fact that the fovea is the center of best vision, the poor peripheral vision of humans arises, meaning that we only see a small part of the visual field crystal clear. Thus, bad human vision starts already here with a good image but a poor resampling due to the fact of the fovea being the point of best vision.

Basic Retinal Circuitry The diversity of cells in the retina is quite large. This is supported by the fact, that the retina contains all neurotransmitters known. The organisation of the retina is so, that the receptor terminals are at the back of the retina and the light has to pass all the layers of ganglion cells etc. beforehand. Therefore, the cells before are transparent. The Photoreceptors are most active in the dark and least active in the light.
As light is a noisy object (it follows in the intensities a Poisson distribution). The cells in the retina are as well responsible for a lot of noise like the transduction noise, channel noise, synaptic noise etc. The entire system is therefore very noisy but as an entity very robust.

Lateral Geniculate Nucleus The *LGN* is organised in layers that are called the *parvocellular layers*. The layers between are called the *coniocellular layers*. The Layers below that are a bit larger are called *magnocellular layers*.

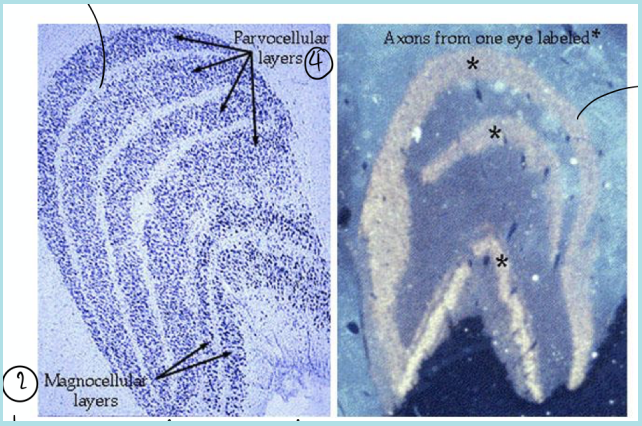


Figure 1: The LGN has several layers, the parvocellular layer (4), the magnocellular(2) which is a bit smaller and the coniocellular in between the other layers.

The brains of mammals can be organised quite differently, with the visual cortex being once in the front, larger or smaller. As the name suggest it is part of the cortex, so the outer layer of the brain. In animals with poor visual processing, the visual cortex can be smaller.

Retinotopy *Retinotopic* organisation means that the same order is kept in following retinotopic compartments as in the retina. That means things that were close together in the retina will remain close together. Examples for retinotopic tissues are the LGN and the primary visual cortex. The fact that there is processing doesn't alter this organisation. The same thing applies to the auditive system, there it is called a tonotopic organisation.

1.1 Striate Cortex

Striate Cortex *Striate cortex* is another name for the primary visual cortex or V1. A characteristic of the striate cortex is the layer 4 which spans it. The striate cortex receives the sensory inputs of the thalamus. The striate cortex has 6 layers.

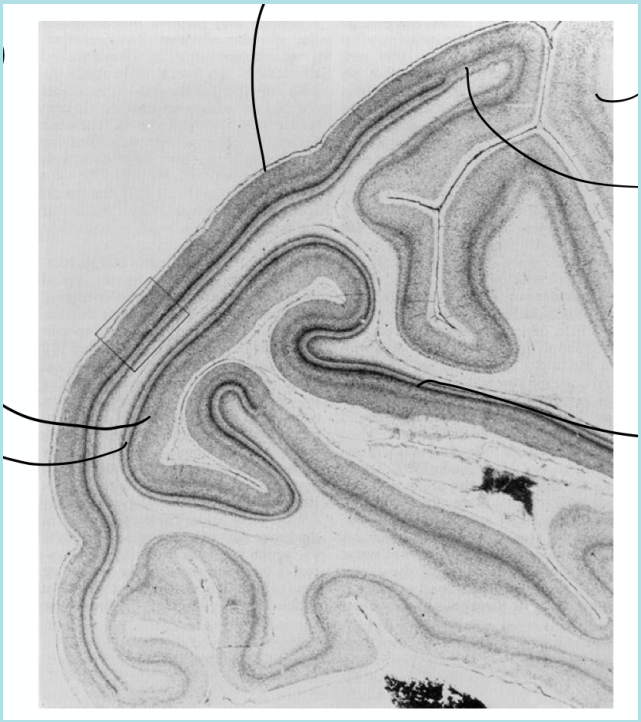


Figure 2: The layer 4 is visible in a dark band of cells. Where the layer 4 ends, ends as well the primary visual cortex.

If one performs a tracer experiment and illuminates only one part of an eye you see that only certain cells have the tracer enriched. These are the *occular dominance* columns (OD columns). These columns receive the same information from one eye and are ordered in columns of cells. From the top this looks like stripes - ocular dominance stripes. The cells above the Layer 4 are binocular, whereas the layer four cells are organised in occuluar dominance columns and are monocular, so that only the information of one eye is processed.
If one colours the regions in the cortex responsible for vision, one sees that $\approx 50\%$ is used for vision.
When looking at the wiring of the cells in the visual pathway we see that there is a lot of feedback. Underlying is a hierarchical structure but there is nearly always some sort of feedback, even from the LGN. When showing the relative importance of the tissues and their wiring we see that the V1 and V2 of the cortex are very important and their wiring between as well. The cortex is differentiated due to morphological reasons, different decoding and "hosting" different parts of the visual field.

1.2 Extrastriate Cortex

Extrastriate Cortex The *extrastriate cortex* is everything outside of the V1, so the striate cortex.

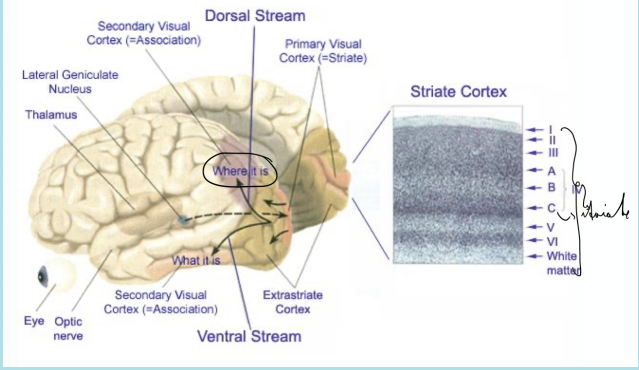


Figure 3: In this picture the different locations of the striate and the extrastriate cortex are visible. The striate cortex (V1) is further subdivided in several (6) sublayers. Note the dorsal stream that is important for landmark discrimination and the ventral stream which is significant for object discrimination.

Object and Landmark Discrimination are two features that were tested with lesions in monkey brains. Doing cross-experiments the regions important for object and landmark discrimination could be identified. Whenever one of them was destroyed, the monkey could only continue performing the other task. The regions, that were affected by the lesion experiments, are found in figure 3.

1.3 Receptive Field

The Receptive Field of a cell is the region of visual space in which light can affect (increase or decrease) the cell's firing frequency. The receptive field depends on the point of view and the scale. The receptive fields of the retina are round and have off surrounds and on centers (or vice versa). In RFs the ganglion cells might overlap so that a change from certain parts onto other parts of the receptive fields can't be visualised. In the middle of the focus the receptive fields are usually smaller and more dense than in the periphery. If a spot of light hits an on center, the ganglion cell fires more action potentials, whereas if the off-surround is hit, the firing of action potentials is lower than usual of the ganglion cell.

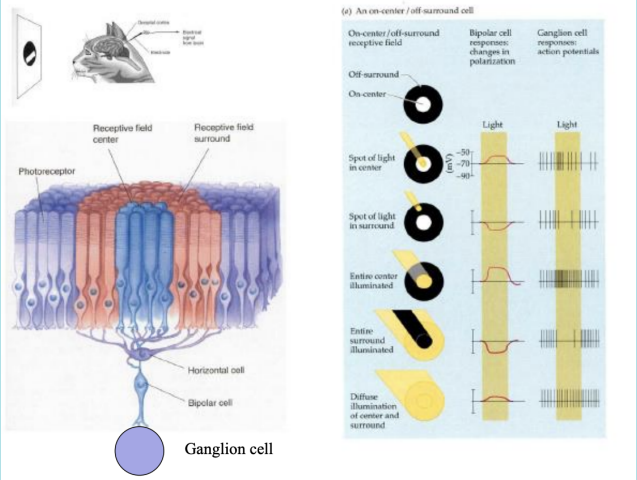


Figure 4: The receptive field of a ganglion cell with an off-surround and an on-center.

Cortical receptive fields can have receptive fields that overlap and work as an entity. These receptive fields can be orientation-sensitive meaning that they respond strongly to a stimulus of e.g. a vertical stripe. These cells can as well be direction selective. The experiments can be made intracellularly by inserting an electrode into the cell of interest or extracellularly by putting a voltage amplifier in the vicinity of a cell.

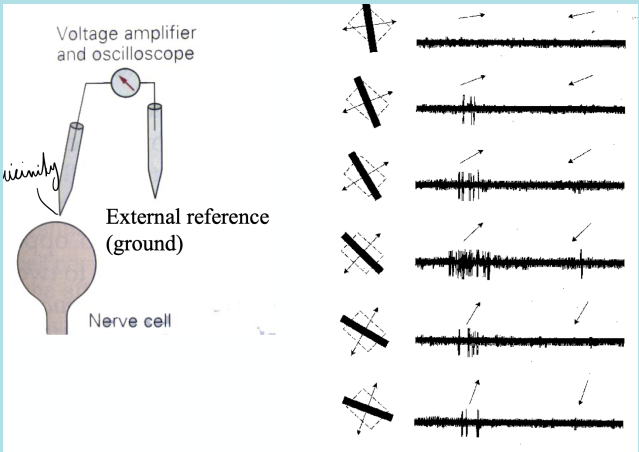


Figure 5: Example of the extracellular recording which show stimulus orientation and direction in the striate cortex.

1.4 Reading Visual System - Wikipedia

Most of the optical neurons end in the LGN from where they are transduced to the V1. The V1 performs some edge-detection to understand the spatial organisation Some pulses of the LGN are as well transduced to the V2 which either forwards them to the V1 or processes them. It is very similar to the V1, an additional capacity being depth-perception via illusory contours. V3 is responsible for global motion like direction or speed and V4 recognises simple shapes. V5 does the same as the other V's, just that it integrates local and global motion.

Mechanisms of generating visual signals In the dark the retinal of rods is in *cis*. When encountering light the retinal changes to *trans* and breaks away from the opsin. This breaking away blocks the production of inhibitory glutamate so that the bipolar cell fires.

Optic nerve About 90% of optic nerves go to the LGN, coming from different parts of the retina. This parallel processing, meaning that each information goes through a different path is important for visual reconstruction. At the optic chiasm the field of view of the right side crosses sides to be processed in the left half of the primary visual cortex and vice versa.

LGN The lateral geniculate nucleus is made up of 6 layers in humans and is found in the thalamus. Layers 1, 4 and 6 are from contralateral (crossed) nerves from the nasal visual field and layers 2, 3 and 5 from ipsilateral (uncrossed) fibers. There are magnocellular and parvocellular regions.

Visual Cortex The part that receives directly information from the LGN is called the primary visual cortex. Visual input by attention starts at the V1 and flows then through a cortical hierarchy; V2, V3 etc. Neurons in V1 and V2 react selectively to bars of specific orientations.

2 Retina

Rod and cone cells Generally, *Rods* are used in low-light (night) vision, whereas *Cones* are used for bright, day-light stronger light. Both types of receptor can adapt well in their favoured conditions. The distribution of cones is centered around the *fovea* and decreases towards the periphery. The *fovea* is an extremely cone rich area, where images are focused and most clear in daylight conditions. Primates have three types of cones, green-sensitives, blue-sensitives and red-sensitives that determine colour vision. Sensitivity refers here to preference of a certain wavelength. Colour blindness is the result of the absence of one of the cone types. Even though rods are also wavelength selective the *principle of univariance* explains why we cannot perceive colour with rods.

The principle of univariance states that the firing rate of cells depends only upon the number of photons absorbed. The responses of receptor types are determined by the likelihoods that their respective photoreceptor protein absorbs photons of different wavelengths (i.e color). A specific signal (firing-rate) can be produced by various inputs, only the combination of signals of the different colour sensitive cones determine the true colour. As there is only on type of rod cells, colour vision in dark conditions is not possible for humans.

Light and dark adaptation The visual system is highly adapted to contrast and contrast levels. The higher the background light intensity, the higher the threshold of activation. This is achieved in two steps, first the rods adapt in darker environments, following by the cones adapting. This two step process can be clearly seen in a diagram. The adaption to darkness takes time in which the photoreceptors have accumulate their photosensitive molecules.

Contrast is proportional to reflectance. The visual system is well adapted to surface reflectance and not to the light intensity. What matters most to our visual system is the *Local contrast* c. Given the Intensity $I = r \cdot i$, which is determined by the illuminance i times the reflectance r, the local contrast is

$$c = (I - \text{mean}(I))/\text{mean}(I).$$

This system is responsible that we do not notice gradual changes in lighting as the local contrast stays the same in different lighting conditions. However, this also lead to a light adaption that is somewhat local in space. As a result of this we tend to perceive the same colour lighter or darker depending on the background lighting intensity.

2.1 Basic retinal circuitry

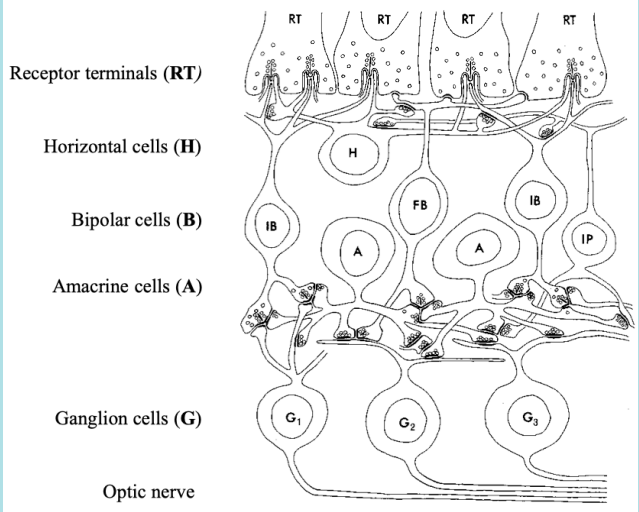


Figure 6: Note that the light enters at the bottom of the image and passes through the retinal layers before it reaches the receptor cells.

Light is refracted by the *cornea* and the *lens* and is projected inverted onto the retina. There it transverses the retinal layers before reaching the receptor cells. There the light, specifically photons, is absorbed by one of the two types of *opsins* (i.e. rod and cone opsins). In rod cells the energy of the photon changes the conformation of the *retinal* from *cis* to *trans* which in turn breaks off the opsin and inhibits the release of neurotransmitters from the *bipolar* to the *ganglion cells*. In the presence of light the *glutamate* secretion is stopped and the cells go into a hyperpolarized state which in turn stops the inhibition of signalling form the bipolar cell to the ganglion cells. The ganglion cells collect information and conduct action potentials to the brain. This process results in the creation of center-surround *receptive fields*. Additionally, *Horizontal* and *amacrine cells* transmit information laterally in the same layer creating complex receptive fields, i.e. either indifferent to color and sensitive to motion or sensitive to color and indifferent to motion.

ON and OFF pathways Different *bipolar cells* contain different types of receptors for the neurotransmitter *glutamate* leading to different responses to stimuli. Some bipolar cells show faster response other slower response to the same amount of stimulation. The response to glutamate in such manner is called the *OFF pathway* detecting dark images against a lighter background. Other bipolar cells have inhibitory glutamate receptors (prevent bipolar cell from firing) which activate the *ON pathway* detecting light images against a darker background. This is only the case in *cone* cells. In *rods* however, the bipolar cells transmit an ON signal and with use of different amacrine cells convey information to ganglion cells. This system also uses information from the cone bipolar-to-ganglion pathway and therefore the rod pathway is very well adapted to twilight and night vision.

Receptive Field Physiology Receptive fields vary in their size. Horizontal cells tend to have large receptive fields as they collect input from many cones and the plasma mebranes of neighbouring horizontal cells fuse at gap junctions. A single bipolar cell receives input from only a handful of cones creating a medium-sized receptive field. The combination of a single bipolar cells with an ON or OFF response and an opposing signal from a horizontal cell forms a *center surround* organization (Figure 4). Horizontal cells can modulate photoreceptor signalling under different lighting conditions as well as shaping the receptive field of bipolar cells and making the bipolar cells response color-coded through feedback circuits to the cones.

Ganglion Receptive Fields Ganglion cells gave receptive field organized as concentric circles. Amacrine cells form circuits that convey additional information to the ganglion cells likely sharpening boundaries between receptive fields. The two basic types of ganglion RF, ON and OFF center, are described in section section 1.3. The fovea contains a special type of *midget* ganglion cells that channel information from a single cone to the brain. The center surround organization is special to humans/primates and can differ in other animals. The ganglion cells are capable of adapting to the mean light intensity. In addition, the center-surround organisation enables the receptive field to enhance edges. Depending on the size of the receptive

field the resolution of this edge detection is more or less sharp.

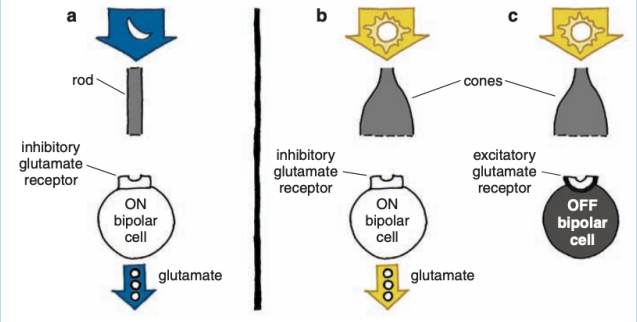


Figure 7: Counterintuitively, photorecptors stop releasing glutamate when exposed to light, but this in turn causes ON bipolar cells to release glutamate.

2.2 The linear model of receptive fields

If a system qualifies as a *linear system*, it is possible to use the response to a small set of inputs to predict the response to any possible input. The following conditions determine whether a system ca be classified as linear:

- *Homogeneity:* $L(ax) = aL(x)$ This means that the response of a system scales with the input. If we double the input we expect the response doubled as well.
- *Additivity:* $L(x+y) = L(x)+L(y)$ If a system is linear, the measured response is the sum of the responses to the individual inputs.
- *Superposition:* System that satisfy homogeneity and additivity are *linear systems*. The combination of these two principles is commonly referred to as the principle of superposition.

Shift-invariance: The response to a stimulus is the same when presented at different time points, only shifted in time. This is not a necessary condition for linear systems!

Convolution generally describes a mathematical operation performed on two functions, where the output is a third function $(R(x,y))$ describing how the shape of one function $(I(x,y))$ is modified by the other$(F(u,v))$. The term convolution describes both the process of compution as well as the output of the operation. It is defined as the integral of the product of the two functions after one is reversed and shifted. And the integral is evaluated for all values of shift, producing the convolution function.

$$R(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(u,v)\cdot I(x+u,y+v)du\,dv \qquad (1)$$

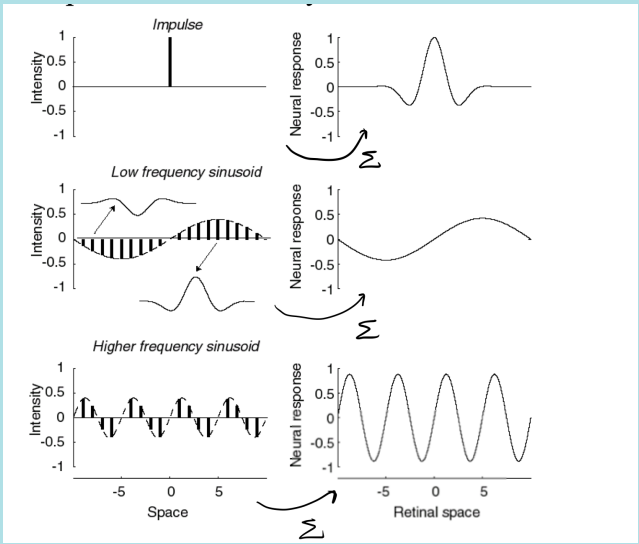


Figure 8: Linearity is often checked by using sinusoidal stimuli, as it holds that: 1) the response to sinusoids are sinusoids and 2) the dependence of response on stimulus frequency can be predicted from the shape of the receptive field. The frequency f is the same, the amplitude A can change.

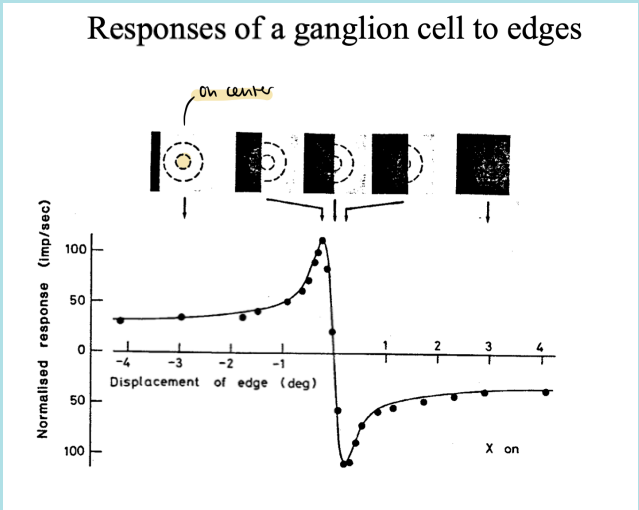


Figure 9: The response of a ganglion cell to edges

The receptive field of ganglion cells can be modelled by a linear model using a *difference of gaussians* model. Both the sensitivity of the center and surround can be described by a distribution that follows a Gaussian shape. The output of the neuron can be described in this model by summing together the responses of the center and the surround. The model assumes that the temporal response of the surround is lower compared to the center. This model can be used to describe the response of a ganglion cell to edges. If we have an ON center ganglion cell and we move an edge over the receptive field we can see that the response increases due to the OFF center being covered with the edge. As soon as the edge passes the middle of the ON center the counteraction of the ON center kicks in and

the response drastically decreases leading to a mirrored image of the response (Figure 9). This response leads to an optical illusion that can be seen in *Mach bands*, where we perceive a change of illumination before the color of a band changes.

Spatial Contrast sensitivity describes the sensitivity to spatial frequency combined with the contrast of the stimuli. It can be seen that people are not equally sensitive to all spatial-frequency patterns, the threshold contrast depends upon the pattern's spatial frequency.

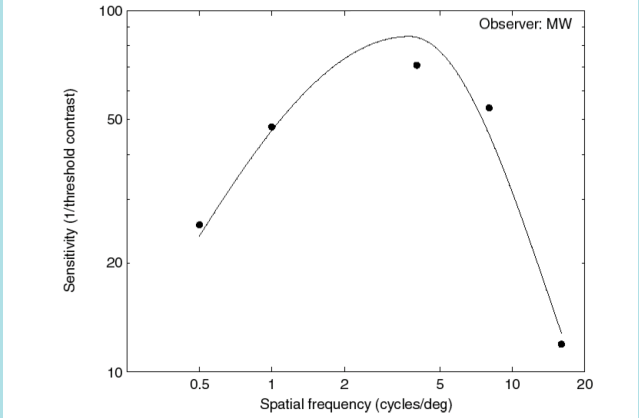


Figure 10: The spatial frequency versus the sensitivity plot. Note that the sensitivity is the inverse of the threshold contrast. This curve is the sum of the individual sensitivities as seen in figure 12.

3 Visual Cortex

The visual cortex can be found in the back of the brain and is differentiated into the striate cortex (primary visual cortex) and the extrastriate cortex. Next to V1, there are around 20 other regions that are known to respond to visual stimuli.
A receptive field is a difference of Gaussians. This behaviour an be plotted as a contrast sensitivity function. The functions narrowness depends on the number of subregions of the receptive field and corresponds to the sharpness of tuning. With more subregions the function becomes more narrow.

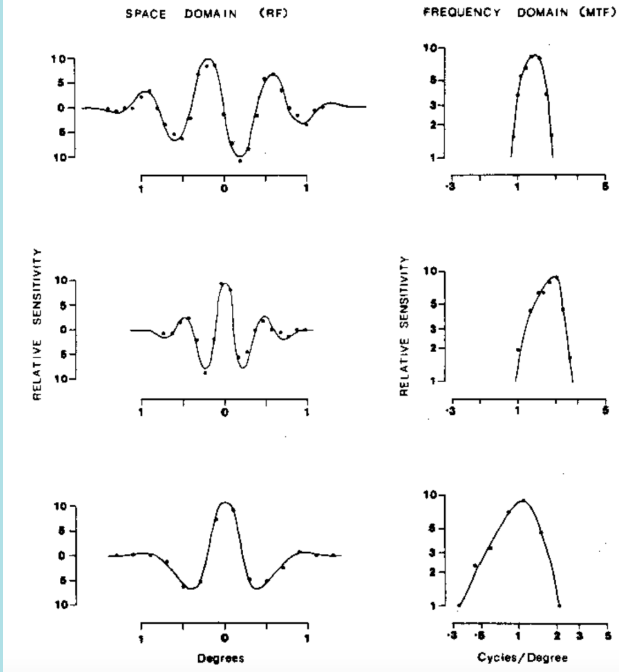


Figure 11: The differences of gaussians leads to the shape of function. The narrowness depends on the number of subregions the receptive field has.

The cortical receptive fields can have the same shape as the receptive fields as in the retina. The frequency domain models the perceptual sensitivity and can be seen as a sum of gaussians that cover like this different spatial frequencies. Measured were the cortical sensitivities depending on different frequencies. In the sum of the differnet cortical RFs the visual field is covered.

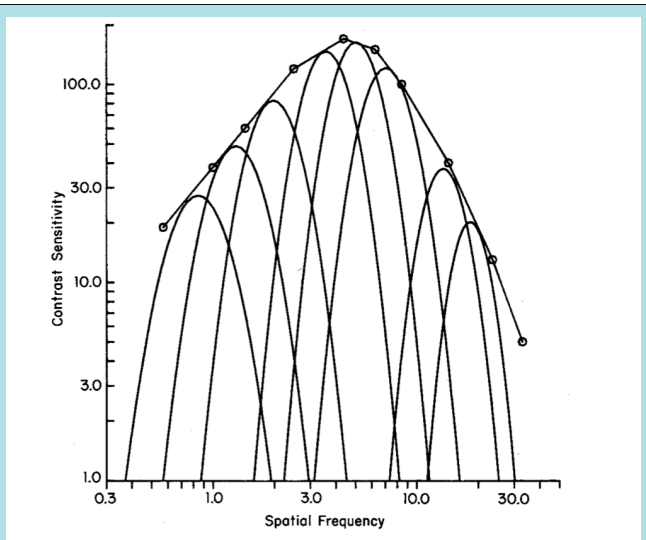


Figure 12: The sum of individual cortical RFs covers the entire visual field of the monkey in this case. Plotted are the sensitivities of the RFs.

The cortical cells are selective for stimulus orientation and direction. Some cells respond well if a stimulus moves in one orientation and direction and poorly if from another orientation and direction. Since cortical neurons

show always a low basal level of activation, one can not say whether the low activity when encountering a stimulus from the not-preferred direction is due to an inhibition or just no activation. This behaviour is seen in figure 5. The behaviour of direction and orientation selectivity can be found in cortical cells but not so in retinal cells or the LGN!
Selectivity in V1 cells is very sharp, meaning that already a small difference in angle can mean that a certain cortical cell shows reduced activity. This means as well that it is very difficult to get an activity of a single cortical neuron.

Retinotopy is a characteristic of the primary visual cortex. It means that the order on the retina is preserved until the primary visual cortex. It is as well preserved prior in the LGN. The receptive fields centers of neurons are aligned systematically from the fovea to the periphery, the only disturbance being the boarders of ocular dominance columns. This can be measured with electrodes or via 2-deoxy-glucose marker experiments. Active neural cells take up the radioactive glucose and the pattern of the visual field is visible in the tissue of the primary visual cortex. The image is though distorted which is known as cortical magnification, since more cortical cells are used to interpret the information of the fovea. This can be seen in plot 13.

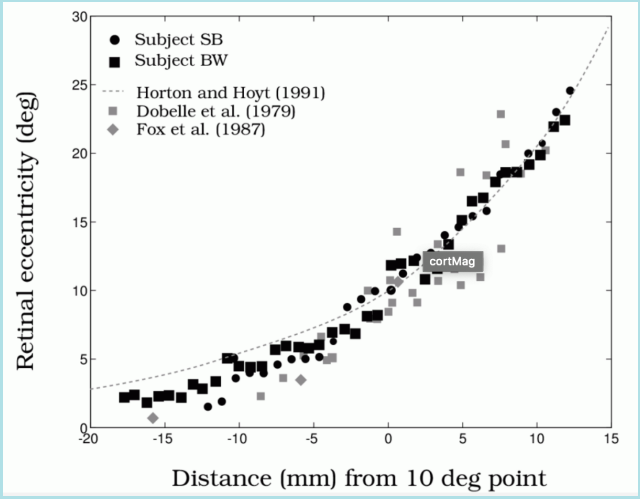


Figure 13: In this picture the cortical magnification can be seen. Plotted are the visual field eccentricity against the location in the calcarine sulcus. The foveal ganglion cells use up more space in the cortex than peripheral ganglion cells. This was measured using fMRI, PET, or microstimulation in blind volunteers.

Orientation preference When plotting the preferred orientation of cells against the track distance one sees that prependicular cells have the same orientation preference. This is similar to the ocular dominance, just in another dimension. Neighbours show a similar orientation preference. One goes tangential through an ocular dominance column and receives these orientation preference regions. These regions are dominated by the contralateal eye (eye that is on the other side as the object in question) and when getting deeper in the ocular dominance column it is dominated by the ipsilateral eye (on the same side as the object in question). The lines of the orientation columns converge into to singular points near the center of ocular dominance columns.

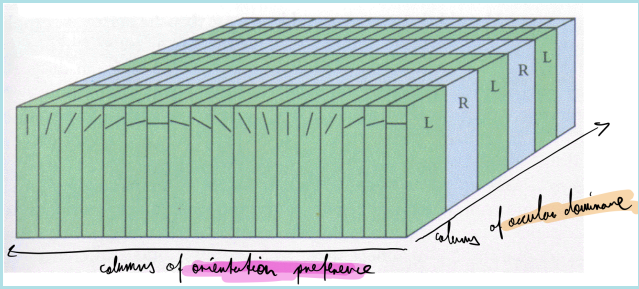


Figure 14: The orientation preference regions are sorted along an ocular dominance column. This schematic neglects the fact that the orientation regions converge into a singular point in the center of the ocular dominance column.

3.1 Linear model of V1 simple cells

"Responses are a weighted average of the stimulus intensity, where the receptive field is the map of the weights."
The linear model describes the response of a cell to a stimulus as a summation that takes signals into account which pass a certain threshold. Depending whether an on- or an off-region records a passing stimulus the cell reacts. The reaction is better if the stimulus fits well in size and orientation to the receptive field. The direction plays as well an important role in the alignment process of the receptive field to the stimulus. It is most strong if the stimulus is of same size as the receptive field and has the same orientation. This in consequence alters as well the spatial frequency tuning of the cortical cell.

Nonlinearities in V1 responses . The linear model obeys to the rules set up in the section 2.2. A basic nonlinearity is thresholding. If we apply a filter to the signal we look at the actual output is not as we would expect with a linear model.
Another violation of linear systems is within the principle of homogeneity via saturation. The observation is that the response of cortical cells doesn't always double if the contrast is increased. The non-linearity arises due to refractory periods etc., which limit the possibility of linear responses. It isn't just due to the number of action potentials as the saturation happens no matter whether the responses are stronger or weaker, meaning it is not just a biophysical limitation.
THe next violation lies in the principle of superposition (section 2.2). The mechanism works as follows namely that there is a mask that is sub-optimal in its e.g. orientation. This cell represses the signal from the optimal cell and the condition of superposition is violated. The

strength of the mask can vary and so does in consequence the strength of the repression.
These factors that inhibit the linearity of V1 simple cells lead to the formation of a nonlinear model. This has thresholding of signals via vicinity cells and intracortical inhibition via masking.

3.2 Adaptation

Adaptation is a change of the system due to adapting to certain environment. This is as such a non-linear process, an example is a so called motion after effect. Adaptation typically happens in the cortex as in the case to contrast adaptation. If you stimulate a cortical neuron with a low contrast, then with a high contrast and then again with a low contrast from the beginning, the average rate of discharging impulses is lower than before - the neuron adapted to the high contrast and reduced it's firing rate after this adaptation. This can be performed the other way around with adapting to a low contrast and then being presented again with a high contrast and recording the difference in e.g. firing rate before and after. This can be seen in picture 15.

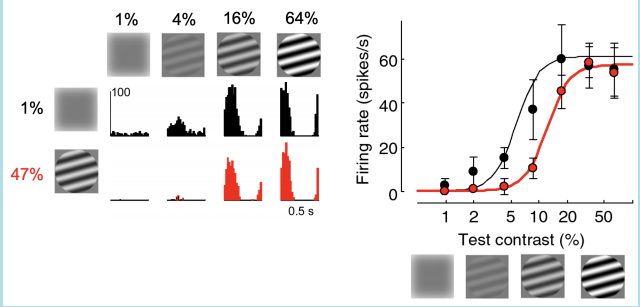


Figure 15: Cells have differential sensitivity when being first exposed to low contrast (black) and showing then high sensitivity or being first exposed to high contrast (red) and then showing less sensitivity to increasing levels of contrast.

3.3 Learning

Learning is again itself a non-linear process as in the linear setting a cell should always react in the same way to a stimulus presented.

Hebbian learning is a way to describe the process of altering the weights of connections between neurons in a neural network. The weight change is proportional to the product of activation values for the individual neurons.

$$\Delta w_{ij} \propto a_i \cdot a_j \tag{2}$$

which is the same as formulating this with a learning rate η.

$$\Delta w_{ij} = \eta \cdot a_i \cdot a_j \tag{3}$$

What this means is that the more often a neuron a_i and a neuron a_j are activated together, the more favorable the two neurons will react to each other. Donald Hebb showed this with differences in synaptic transmission between neurons. This is found to be the basis of the model of synaptic plasticity.

Evidence for Hebbian Learning in V1 are measuring responses of neural activity before and after conditioning to a certain type of stimulus direction. The selectivity of a cell can be changed via an additional electrode. In effect the selectivity changed temporarily.

3.4 Cells of the V1

V1 simple cells are similar to retina cells with clean on / off regions. These can be radial or stripe-wise. On, off-regions can react linearly and the RF field is linear to these regions. This can be tested via reverse correlation. This is a technique that is used to study how they sum the stimuli from different locations together.

V1 complex cells are cells that are still orientation selective. There, the entire region reacts to both positive or negative stimuli. The RF looks more like a mixture of on and off regions, so that there is no clear distinction. This behaviour is not seen in retina or the LGN.
Simple cells are most often found in the layer 4. The LGN sends their axons into layer 4.

4 Motion Processing

The perception of motion is a visual inference. Our brain interprets motion based on two dimensional images we see. Notably, perceived velocity of an object is color and contrast dependent. This response to *moving stimuli* is performed in the *area MT* of the *visual cortex*.

4.1 Reichardt detectors

are hypothetical neural circuits that can describe how the brain is able to track motion and speed. A cell recieves input from two other cells X and Y, the signal of one cell X is delayed such that the signal from the two inputs do not reach the cell at the same time. The cell is activated if both signals reach the cell at the same time, which is only possible if the signal is moving in the correct direction with the right speed. (Figure 16)

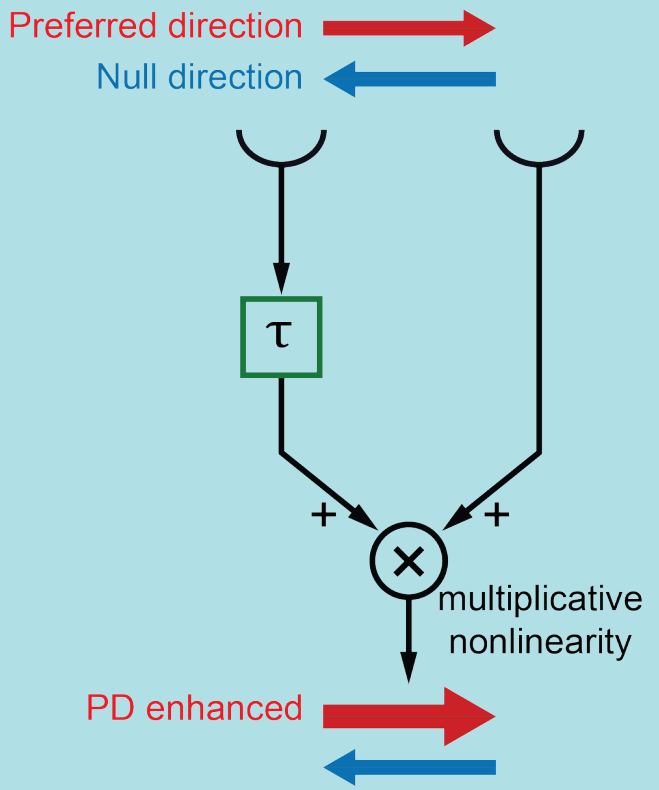


Figure 16: Schematic drawing of the Reichardt detector, where τ denotes the delay module.

4.2 Space-time Receptive fields

Figure 17 shows how we can describe motion. If we have one dimensional motion (e.g. line moving to the x axis) we encounter all relevant information in the (x,t) plot. The slope of the bar in the (t,x) plot refers to the speed of the line. The steeper, the faster the line moves on the x axis.

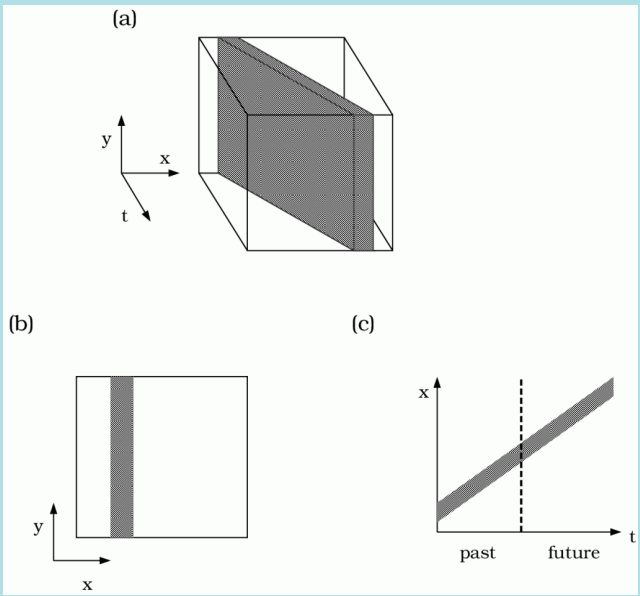


Figure 17: The motion of an object can be represented in (x,y) and (t,x) plots.

We can use this model to study space-time-oriented receptive fields. Figure 18 shows how receptive fields can be tuned to be optimally space-time-orientation selective. The neuron always responds to events in the recent past, so the receptive field moves along the time axis with the present. The common orientation of the space-time receptive field and the stimulus motion produce a large amplitude response. This receptive field can be created by combining the responses of neurons in V1 of the cortex.

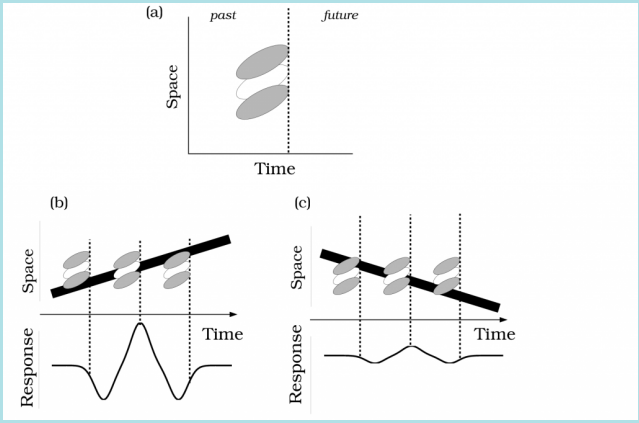


Figure 18: Space-time oriented receptive field

Reverse correlation can be used to study the space-time-oriented receptive fields. It is a tool for characterizing the response properties of a neuron using the spikes emitted in response to a time-varying stimulus. Mathematically, it is the average stimulus preceding a spike. To compute the it, the stimulus in the time window preceding each spike is extracted, and the resulting (spike-triggered) stimuli are averaged.

4.3 The aperture problem

The fact that we can only measure the motion component perpendicular to a constant one dimensional motion and the inability to measure motion along the constant spatial dimension is known as the *aperture problem*. This problem is illustrated in figure 19, where for one RF the perceived motion shows no difference if the square moves to the right or the bottom of the image.

The intersection of constraints offers a solution to the aperture problem. This method uses only the local motions of two edge pieces to compute the global motion by finding the intersection of all possible global motions consistent with the two local motions detected. Note that this is not simply a vector addition. This is also illustrated in figure 19.

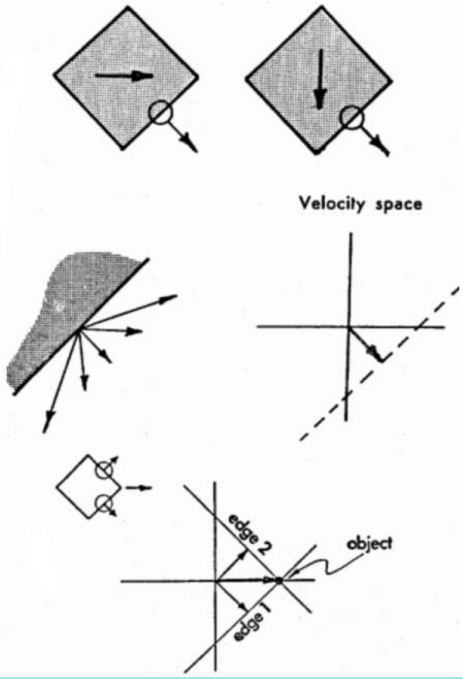


Figure 19: The aperture problem and the solution to it by the intersection of constraints.

Component and pattern direction selectivity Two translating *gratings* (just parallel lines) form a *plaid* (looks like a grid) stimulus. Each grating by itself contains only a single orientation and its motion can not be estimated uniquely. However, when both gratings are added, subjects often assign a single motion to the pattern. This is also done by the intersection of constraints of the two gratings. This can be seen in figure 20 where the response of the same neuron to a grating and plaid stimulus is shown. In the right part of the figure the dotted triangle shows the direction in which the plaid moves, the solid lined triangles show the response of the neuron. Whenever a plaid moves in the direction of the dotted triangle, one component moves in the direction that is indicated with the solid line and therefore we see this type of response. Figure 20 shows a conceptualized response which can be confirmed in *V1* neurons.

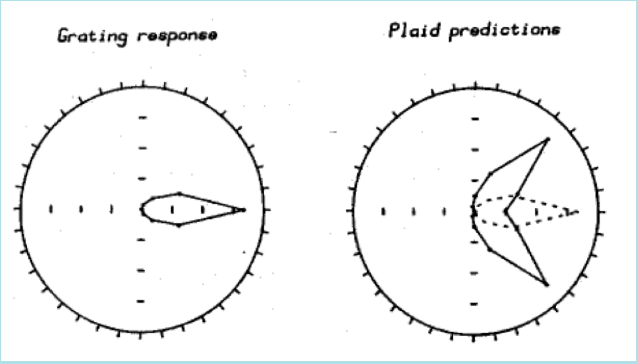


Figure 20: The selectivity of a single neuron to a grating and plaid stimuluts. The units on the circle denote the angle of a cue, the units within the circle the strength of the neural response.

MT cell responses Analysis of the population of cells in the *medial temporal* (MT) region shows that there exist at least two types of cells. The *component selective* cells are tightly tuned to a specific direction (as in 20). *Patter selective* cells are not as direction selective and show that they selective to a grating and a plaid moving in the same direction (even though the components of the plaid do not move in this direction. This is a special computation of the cells that is only observed in MT cells (not in V1 cells). When comparing the population of cells in V1 and MT, we can see that V1 shows correlation mainly with component selective cells, whereas in MT the exists both component and pattern selective cells.

Motion sensitivity can be measured using cues moving in the different direction. If all cues move in the same direction we have 100% correlation, if all move randomly 0% correlation. When measuring motion sensitivity in macaques we can see that under 10% correlation the monkey is mostly guessing, if we have more than 40% the prediction is mostly correct. There exists a very simple model that can describe the motion sensitivity. Let's suppose there exists a *neuron* and a *antineuron*, if the activation of the neuron is larger than the antineuron we predict "right" motion and vice versa. When studying only few MT cells we can see that this model matches with a high correlation. As the are MT is organized in direction selective areas, we can see that if we artificially stimulate the cell in the MT the direction selectivity persists and response is always shifted to the preferred direction.

5 Depth Perception

5.1 Cues to depth perception

- size
- lighting and shadows
- Interposition

- Clarity and elevation

- Perspective

When we look at pictures we usually assume that the light comes from the top. We can usually turn the item and the cue changes.

5.2 Binocular depth cues

Convergence is the process when the eyes move conjunctively which shows that the new object B is closer than point A. This is visible in figure 21.

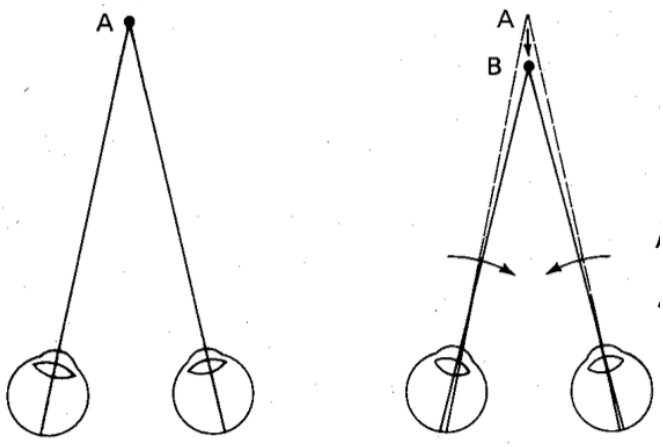


Figure 21: The conjunctive movement of the eyes upon a new object B tells us that the object A is further away than object B.

Binocular disparity describes the term if the difference of the distance away from the point of fixation is not the same. The only objects where there is no disparity upon fixation are such that are on the horopter, meaning that the retinal disparity equals to zero. If there is disparity between the point of fixation and the new focal point, the object is further or nearer than the previous point. This mechanism is described in figure 22.

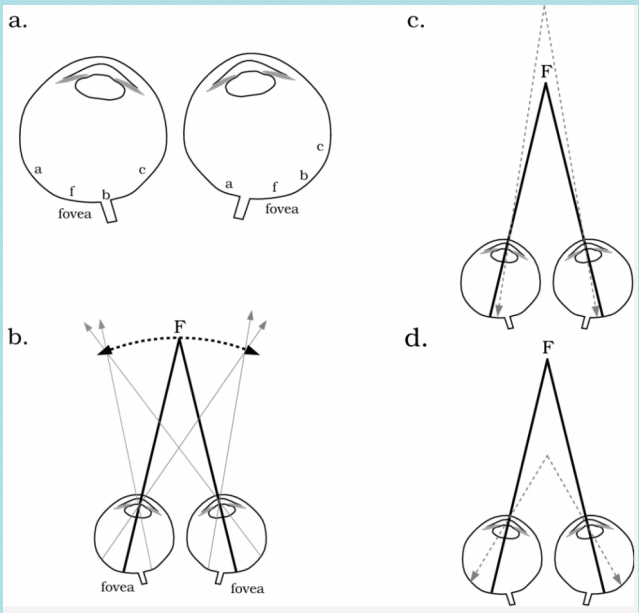


Figure 22: Retinal disparity and the horopter are explained. (a) The fovea and three pairs of points at corresponding retinal locations are shown. (b) When the eyes are fixated at a point F, rays originating at corresponding points on the two retinae and passing through the lens center intersect on the horopter (dashed curve). The images of points located farther (c) or closer (d) than the horopter do not fall at corresponding retinal locations.(From "Foundations of Vision")

Panum's fusion area is an area where two object create a binocular fusion in the eye, resembling to one object, whereas objects liying outside of this area result in diplopia, so seeing two objects. This fusion creates a feeling of visual depth and is employed in 3D movies.

Wheatstone stereoscope was the first 3D generator which allowed two images to be fused and create a feeling of depth. In modern 3D movies there is differentially polarised light and always one eye is "excited". This happens very fast and we see a picture always with one eye and then with the other which creates a feeling of depth as they lie in panum's fusion area and the image occurs to be 3D.

Correspondence problem is the the problem that through geometry several images produce the exact same projection on the retina. How can the eye thus distinguish such objects? The answer lies in the V1. There, some neurons are tuned for a certain type of binocular disparity. The idea is that a pool of neurons only responds to a certain type of disparity, meaning that certain pools respond only to objects being closer than the horopter and others to objects being further away than the horopter. There is some experimental support for this theory in figure 23.

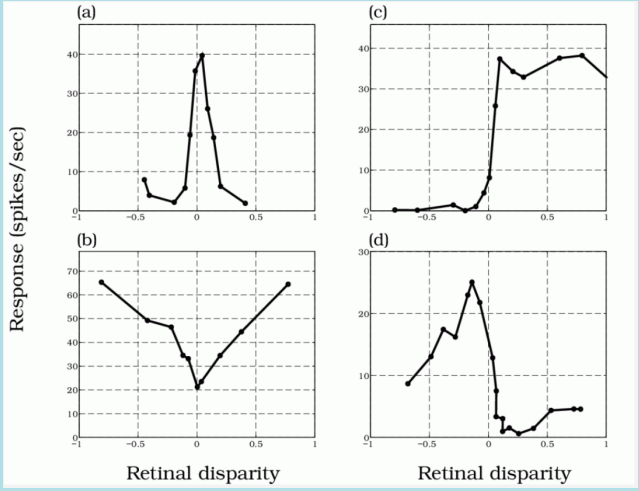


Figure 23: (a) and (b) show the responses of neurons that respond best to stimuli with near zero disparity, that is near the horopter. Responses of a neuron that responds best to stimuli with positive disparity (c) and a neuron with negative disparity (d) are also shown. (From "Foundations of Vision")

The disparity can be seen as well in the receptive fields of neurons. The receptive fields of disparity tuned neurons can seem shifted, where the centers of the receptive fields are e.g. shifted horizontally and show therefore a non-zero disparity.

5.3 Reading Depth Perception

The receptive field (RF) is classically defined as the area of visual space within which the discharge of a neuron can be influenced. The RF is though not only a function of space but as well of time, this description is called the x,y,t plot. The information of neurons in the retina is sequentially processed through the LGN to the primary visual cortex. For retinal and LGN neurons the RF has a circular center-surround shape. There are two sorts observed, an ON-center and another OFF-center version. In the primary visual cortex there are two main types of cells, simple cells and complex cells. Simple cells are the most abundant have spatially organised RFs with alternating elongated subregions that respond to bright or dark stimuli. It is thought that this structure arises from an array of LGN RFs. Complex cells respond to both bright and dark stimuli which are dispersed within the RFs.

Approaches to RF mapping In order to achieve an x-y-t domain plot one typically needs a lot of data, given the stochastic nature of the visual stimulus. In order to cope with the amount of data most techniques sacrifice one dimension. The approach of static RF maps has the disadvantage that spatial and temporal factors are confounded. The response plane technique can create an x-y-t map but it has the disadvantage that it is quite slow. The most efficient startegy is to use pseudo-random spatiotemporal stimuli called "white noise". After the pseudo-random stimulus the neuronal spikes are correlated to the stimulus sequence and like this the neuron's transfer function can be determined.

RF as a spatiotemporal entity RFs are not static but a dynamic entity. There are two main types, space-time separable and space-time inseparable RFs. Space separability means formally that an RF $R(x,y,t)$ can be described as the product of two independent functions, the spatial profile $G(x,y)$ and the temporal file $H(t)$. The RF is then $R(x,y,t) = G(x,y) \cdot H(t)$. If a RF is space-time inseparable, this factorization is not possible. Simple cells range from separable to strongly inseparable. In the latter case, the description of the RF as a static entity is not satisfactory. The two-dimensional spatial envelope of teh RF remains approximately fixed as time progresses, rather the subregions move. This has implications in understanding motion selectivity.

Spatiotemporal RF transformations in the geniculostriate pathway A good way how to show the dynamics of RFs ar x-t plots. For LGN cells an x-t plot typically ehibits a center-surround organisation in space and a biphasic structure in time. There are two main types of LGN neurons, lagged and nonlagged that show different temporal response properties. Lagged cells are distinguished from non-lagged cells in their temporal phase shift. For nonlagged cells the first temporal phase of the RF is largest, whereas for lagged cells the second temporal phase dominates. This accounts for the delay in lagged cells. This behaviour arises in the LGN, as it isn't seen in the retina. Unlike LGN neurons simple cells of the striate cortex show marked space-time inseparability. How these are constructed is not fully understood but one idea is that they are constructed from a pair of separable simple-cell RFs or that they are directly a combination of lagged and nonlagged cells of the LGN.

Spatiotemporal mechanisms that underlie motion selectivity Unlike their geniculate antecendents, most cortical neurons are quite selective for stimulus velocity. Theese are most often found in the striate cortex, LGN neurons seldom show mroe than a weak directional bias. The direction selectivity seams to originate in the linear st st RF structure of simple cells. Most notable, these RFs are tilted(inseparable) whereas st separable RFs are not expected to contribute to direction selectivity. For complex cells that are direction selective these first-order RF profiles don't explain the behavior well enough. In second-order RFs however spatio-temporally oriented RFs can be found.

Origin of ON and OFF responses simple cells and LGN RFs have spatially segregated ON and OFF subregions, whereas complex-cell RFs do not. If a cell's x-t profile shows a bright-excitatory pahse followed by a dark-excitatory phase, this is considered an ON region. If it is the other way around, we speak of an OFF response.

The inaccuracy in these mappings to ON and OFF regions comes from the fact, that in static pictures, space-time inseparable configurations were not taken into account.

6 Colour vision

When white light passes through a dispersive prism it can be seen that it is composed of different colours. Isaac Newton discovered this and the fact that the white colour can be recombined when its components pass through a different prism. In very low light levels, vision is *scotopic*: light is detected by rod cells of the retina. Rods are maximally sensitive to wavelengths near 500 nm, and play little, if any, role in color vision. In brighter light, such as daylight, vision is *photopic*: light is detected by cone cells which are responsible for color vision. Cones are sensitive to a range of wavelengths, but are most sensitive to wavelengths near 555 nm. Between these regions, *mesopic* vision comes into play and both rods and cones provide signals to the retinal ganglion cells.

Real life colour perception Which colour we perceive depends on multiple parameters. The product of the *illumination* that is different for different wavelengths and the *reflectance* of a surface that is fixated gives the *colour signal*. This colour signal is the physical input to our visual system. There, the *cone sensitivites* and the relative *cone absorption* determine what colour we assign to the fixated object. The colour is the reconstructed based on a combination of the signal of the *L, M* and *S* cones.

Photoreceptor sensitivities The different photoreceptors have different sensitivities. This is the case for the *rods* as well as the three different *cone* types. The sensitivity profiles peak at a certain wavelength but show some overlap. Colour reconstruction is based on the reconstruction of the *3 primaries* that are given by the different cone cells, i. e. an additive colour model. Despite the fact that the rods are also wavelength sensitive, colour perception is not possible with rod vision as there is only one type rod cells.

Irregular human colour vision Colour blindness is most often the result of a defect on one of the three cone types. Colour blindness affects mostly men ($\tilde{1}0\%$ of the male population) as it is *X chromosomally* inherited. *Protanope* is a defect in the red (L) cone cells, *deuteranope* is a defect in the green (M) cone cells and *tritanope* a defect in the blue (S) cone cells.
Tertrachromacy refers to a fourth cone type. This is the case for some women that have a fourth cone cell type dure to a mutation. These women are able to see wavelengths beyond those of a typical human's vision, and may be able to distinguish between colors that, to a normal human, appear to be identical.

Cone mosaic When studying the distribution of photoreceptor cells on the retina it can be seen that S-cells are missing from the center of the fovea (as well as rod cells which are completely missing). Comparing different humans it can further be seen that the distribution of the photoreceptor cells varies significantly between different individuals. The ratio of S to L and M cones is constant but that of L to M cones varies a lot. Colour discrimination is the same over all distribution but it remains the question if the subjective colour perception it the same for all people. In philosophy and psychology the instance of a subjective conscious experience of a stimulus is called *qualia*.

Colour matching experiments In colour matching experiments a subject has to adjust the three primary lights until the colour of a test is matched. This is used to characterise human colour vision in a standardised way. The results show the distributions of primary lights that are needed to get a certain colour/wavelength. For some wavelengths there exists a "negative intensity". In these cases the colour cannot be matched with a simple mixture of the primaries but the intensity of the test light has to be altered as well. The colour matching experiments can be very well explained by predictions and is very constant across different subjects.

Metameric spectral distributions describe different distributions of wavelengths that produce the same colour sensations in human observers.

Receptive fields of LGN neurons As previously seen LGN cells can have different receptive fields. We have already seen the receptive fields (RF) of *M cells* which have either *ON-* or *OFF-center* RF based on contrast intensities. But there exist also the *P cells* which have *ON-* or *OFF-center* RF based on wavelength differences. There exists two combinations of wavelength selective receptive fields:

1. red-green receptive fields (either *ON-* or *OFF-center*)
2. blue-yellow receptive fields (either *ON-* or *OFF-center*)

This combination of wavelength selective receptive fields might be the physiological reason for our perception of *antagonistic colours*. These receptive fields that are colour specific receive input from the respective wavelength selective cones. The yellow centers receive a mixture of inputs of the green and red cones as there exist no yellow selective cones.

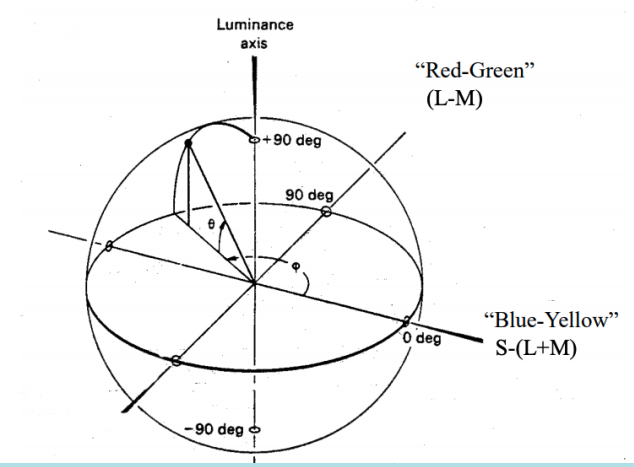


Figure 24: In the DKL colour space each colour can be represented by a specific point in the colour space. The origin of the space is a neutral grey.

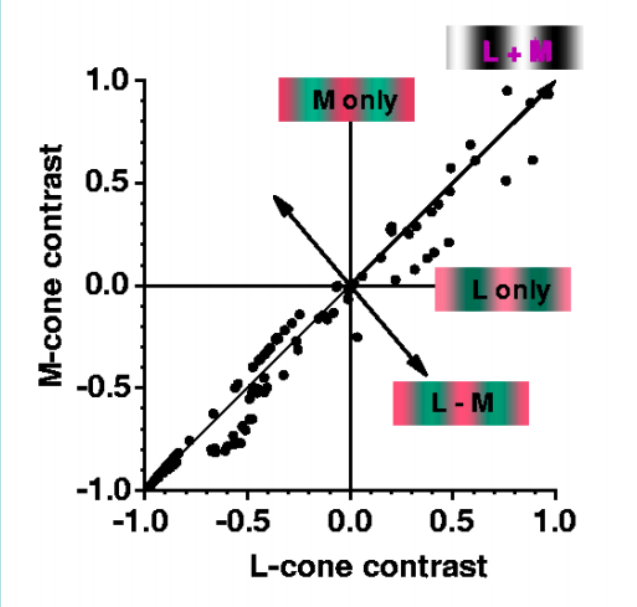


Figure 25: Colour contrasts of natural objects. Each point represents a pixel in the image.

The DKL colour space The dimensions of the DKL colour space represent the following,

- the z-axis encodes for the *luminance*. *Isoluminance* describes colours with same intensity that are on the same vertical plane on the z-axis,
- the x-axis is the *blue-yellow axis*. It is a the difference between the responses of S-cones and a combination of the L and M cones [S - (L+M)],
- the y-axis represents the *red-green axis*. It is the difference between the responses of the L and M cones [L-M].

Each colour can be described by two angles, ϕ the azimuth, which is the bearing in the colour space, and ω the elevation, which is the bearing in the luminance space.

pLGN neuron colour preference Studies show that the neurons in the *parvocellular* layer of the LGN fall neatly into different categories of preferred colours, which are red, green, yellow and blue. Within these categories there are neurons that prefer different elevations or luminances of the specific colour. The LGN therefore splits the information of an image into three different channels: the light intensity channel, the red-green channel and the blue-yellow channel. Figure 25 shows a representative distribution of the excitation of neurons in response to natural images. It can be seen that the neurons respond to either a difference of L+M (difference in luminance) or the L-M (difference in colour), which correspond to two of the channels. This is the ideal way how to respond to colour differences in general and a way in which most colour information can be processed. The two channels are the two principal components of the distribution. Note that in this image the S-cones are missing, but when including the S-cones in the analysis it can be seen that the S-cone would add another channel and principal component to the image.

Cortical cell colour preference Cells in area V2 and V3 (extrastriate cortex) generally prefer a high elevation (high luminance contrast) but also some colour variation. Cells that only care about colour variation are very rare (in contrast to the cells in the LGN). There is no characteristic colour clustering like in the LGN and the distribution is much more uniform. As the differentiation into the channels seems not to be present any more some cells like intermediate colours.
Further analyses of cortical cells show that the responses of cortical cells to colour can be very different. Some cortical cells react to a linear combination of cone signals and have a broad tuning in colour selectivity, whereas other neurons are more tightly tuned to a specific colour. However, there seems not to be many cortical cell that are interested in colour compared to other patterns.

Three stages of colour processing

1. three type of cones: red, blue and green
2. retinal ganglion cells or LGN neurons that react to luminance differences, red-green or blue-yellow variations
3. cortical cells have higher order mechanisms. Cells react to much more variations of colour and can be more tightly tuned to specific colour and more selective to certain colour.

Colour appearance and constancy in later stages Although it is quite clear how colour is processed in early layers it is still very unclear how colour is processed in later stages. For most optical illusion it is still unclear what the physiological explanation for the illusion is. It seems

that colour constancy and appearance depends on luminance and reflectance of an object.

7 Computation in Noisy Systems

Definition of Noise:"Random or irregular fluctuations or disturbances which are not part of a signal(...) or which interfere or obscure a signal or more generally any distortions or additions which interfere with the transfer of information.
In the visual system (and lot's of others) signals are often embedded in noise. Depending on the difference between signal and noise, the signal can be visible/non-visible in the noise it is surrounded by.
There are some key distinctions to be made when considering noise.

- External internal noise
- Internal Noise
  - Cellular noise
  - Electrical noise
  - Synaptic noise
- Motor noise

Noise comes about from the stochastic nature of e.g. signalling, signal propagation in neurons, synaptic vesicles and even in the motor neurons, which show stochastic component in their reaction / flexibility to the stimulus present.

Cellular Noise When looking at the cellular response to neural activity along an axon, we see that the signals are a) shifted in time (show a certain delay) and b) the spread of the signal (uncertainty in the estimate) increases as the signal propagation continues on the axon. This is consistent with the idea, that the error is amplified the further the signal is carried on.
Another example is *network noise*. Noisy elements add up and the integration of these gives an even noisier sum. This is first of all due to the fact of biophysics so the mere physiological component and as well noisy integration, explained above.

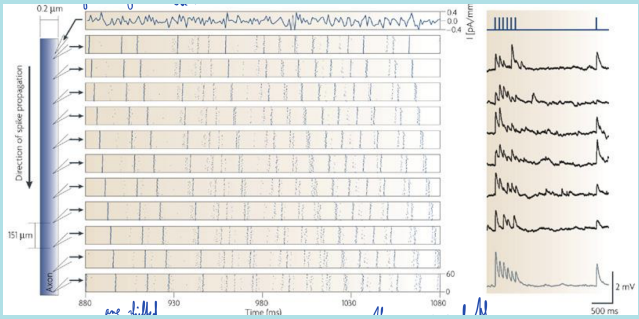


Figure 26: In this picture we see in a) the signal propagation along an axon. Clearly visible is the shift in time with further distance from the starting point and a more pronounced error in the signal as the current continues.

Synaptic Noise An example for synaptic noise are the miniature potentials. These are small signals or excitations that happen at random at the synapse without any specific input. It is due to the stochastic release of neurotransmitters.

7.1 Solutions to noise

The question that remains is now how the nervous sytem deals with noise. There are several different solutions

- Spatial averaging
- Temporal averaging
- Prior knowldege (Bayesian theories)
- Specialized noise reduction

Spatial averaging One solution is to average over space. That means several signals are averaged into one output signal. This averages out the significance of individual errors. An example for this is lateral inhibition. In the retina several photoreceptors are pooled together and excite one ganglion cell. This is the process of pooling over a spatial area and thus spatial averaging.

Temporal averaging is a technique which employs e.g. lowpass filtering. Lowpass filtering means that below a certain frequency there is a cutoff and very low frequencies are filtered out. Like this the noise is filtered out to a certain extent and the curve gets smoothed.

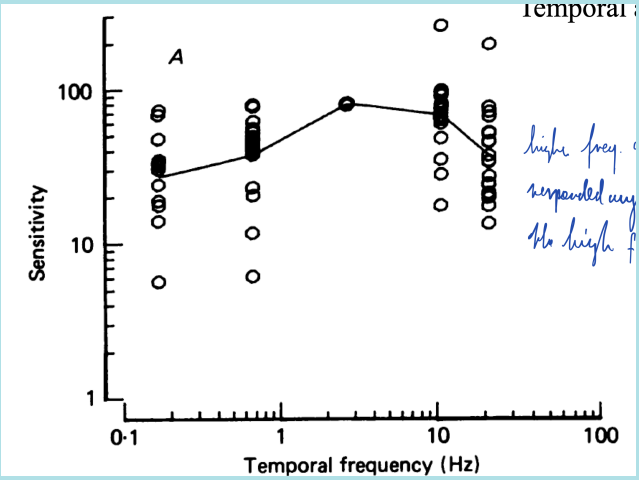


Figure 27: Here the process of temporal averaging is visible. Above a certain frequency the cell doesn't respond anymore and has a sensitivity of the basal state. This can be seen as a process of highpass filtering where frequencies above a certain threshold are filtered out. This results in the curve above.

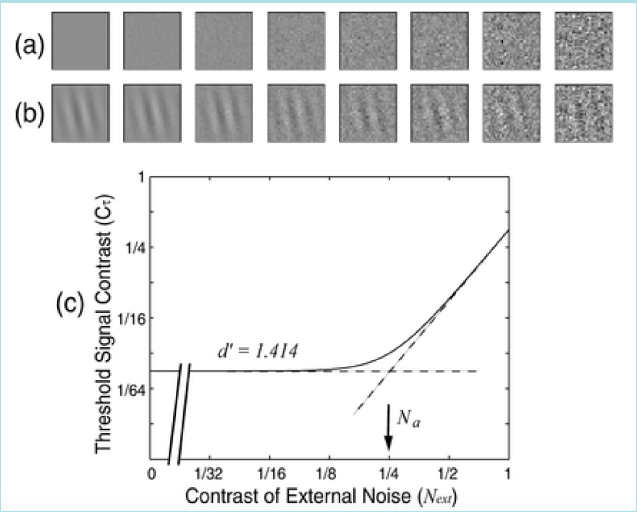


Figure 28: In this experiment Denis Pelli set out to quantify the internal noise in the human visual pathway. By applying increasing levels of noise he measured how much contrast was needed to detect the pattern presented. This was a way how to measure the contribution of internal and external noise, because only at the point of inflection of this curve the external noise becomes more dominant than the internal noise and therefore there internal noise is $\leq$ external noise.

7.2 Noise in vision

There are two important behavioral studies that tried to quantify noise in the visual pathway. **Hecht, Schlaer, and Pirenne** built and apparatus in 1942 which allowed to project very dim and short flashes of light onto a test person. Like this they tried to quantify what was the dimmest light that the human observer could possibly see. What they found was that humans are very sensitive to light and are able to see when several photoreceptors are activated at the same time. Nevertheless the human observers saw random flickering and specks in the dark. This was due to random firing of neurons called *dark noise*. Therefore the actual light for perceiving something was the background light plus the dark noise and together these two things pose the barrier for detection of light in the dark. That explained why the observers needed more photons than just very few.
Denis Pelli set out to measure a different thing namely the internal noise. In order to do that he measured the threshold signal contrast when presenting gradings of contrast with increasing levels of noise. The contrast didn't need to increase before as there the internal noise was the limiting factor, not the external noise. At one point though, the external noise was stronger and became the dominant force and the contrast had to increase. At the point of inflection, the internal noise was $\leq$ the external noise. This was a way how to quantify the magnitude of internal noise in the human visual system.

8 Coding and Decoding of Visual Information

Coding and decoding are two very important processes in all parts of the sensory system. Since the visual system is noisy it is especially interesting. Researchers found that for coding the rate of firing is more important than the exact time of the action potential.

8.1 Noise Correlation

Two neurons can have variable responses. Their noise can be variable as well which leads to different reactions of neurons to stimuli. The question is now whether or not the noises of the neurons are correlated and if they are whether this relationship is positive or negative.

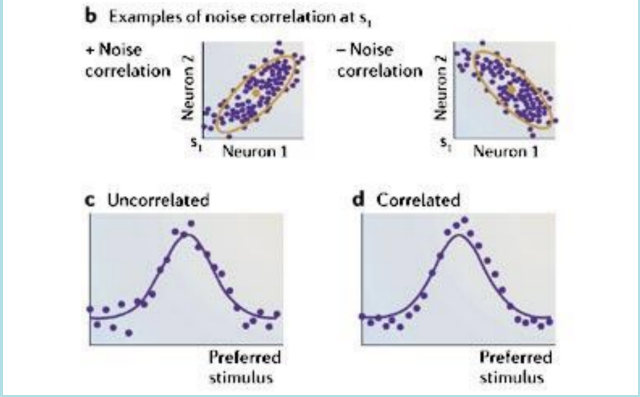
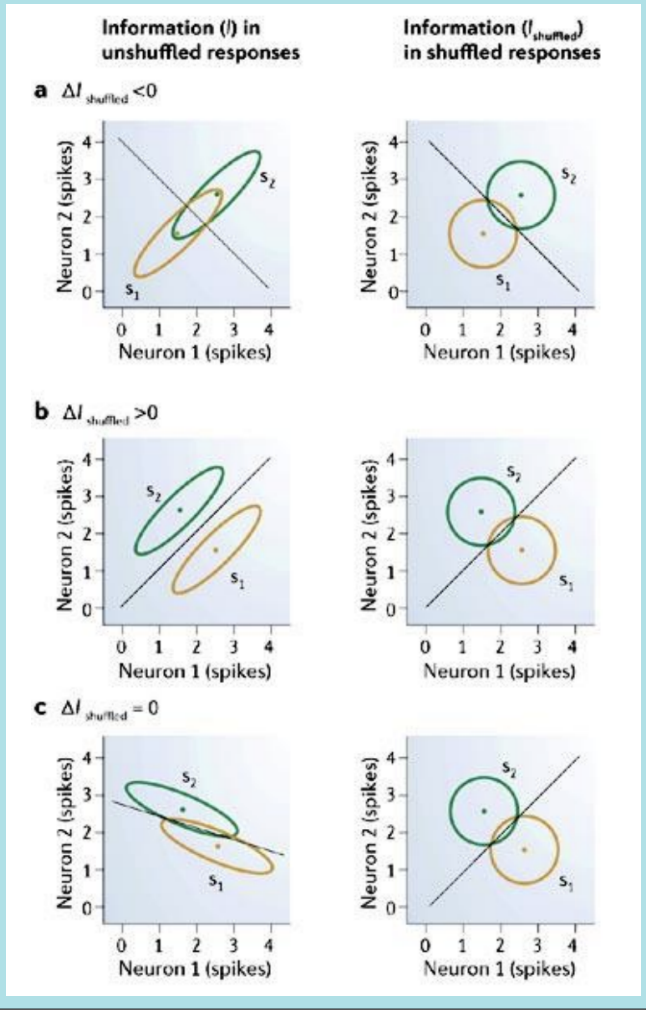


Figure 29: in picture b the examples of positive and negative noise correlation are shown. In picture c the uncorrelated response is shown vs. the correlated response. There we see that in the uncorrelated case neighbours have quite a variable response whereas they show very similar behaviour in the correlated response.



These findings were as well tested using simulations by having a set up with unshuffled response so a clear correlation between neighbours and a shuffled response were there is effectively no correlation visible. In correlated neurons the area of uncertainty might be larger than in uncorrelated neurons as the overlap between the noise regions is larger. If one neuron responds more to one stimulus than to another, the picture is a little bit different as there will be a better separation of the two in the correlated response compared to the uncorrelated response where it is still more or less random where they lie. If the loss of information is more or less 0 there is no clear distinction between the case of correlated and uncorrelated. This is coherent with different depictions showing that the information gain is the largest when applying a weak negative correlation.

8.2 Synchronisation Hypothesis

Researchers were for a long time startled at how the visual system could distinguish between objects. Despite being processed at various areas in the nervous system our percept must recognise what belongs to the same object. This is what is referred to as the binding problem. The idea put forward was the synchronisation of single spikes in the visual pathway. This works by stating that the binding is represented via synchronous rhythmic firing of the neurons selective for the paired features. This joined activation allows the brain to process features together as belong to one object. The firing is synchronous for one object but different between two objects.
This was tested by measuring the reaction of two neuronal areas in the cat cortex upon correlated stimulation. If the stimulus was correlated (moving in the same direction and at the same speed), there was a correlated firing between the two neurons, however if the stimuli were not correlated, meaning that they e.g. moved in different directions, there was no correlated/synchronised response in the neurons recorded. This can be seen in figure 30.
Synchronisation might explain as well the phenomenon of **ambyopia**. This is the abnormal development of the visual system, one example being squinting. If squinting is not treated early on in development, the weaker eye will always perform worse. In the case of synchronisation the information of the squinting eye will be completely disregarded and only the information of the strong eye is used.

As stated before, the most important process for information coding with the least loss of information is rate. The higher the window length (the time) is, the more positive is the influence of the rate until it eventually saturates.

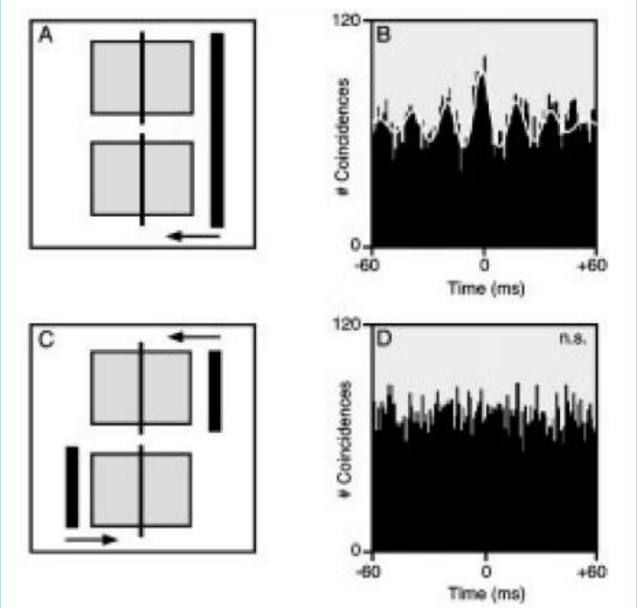


Figure 30: The effect of correlated stimuli on a synchronised/correlated response of two neuronal areas is visible. In the upper picture the response is synchronised, as the stimuli are synchronised but in the lower picture they are not synchronised, leading an asychronous response.

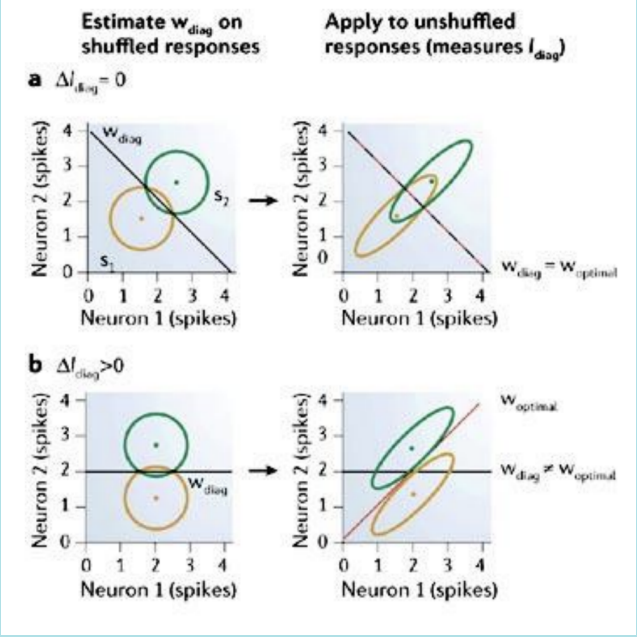


Figure 31: The knowledge of the type of correlation is beneficial depending on the context. If the $\Delta I_{diag} = 0$ $w_{diag} = w_{opt}$ whereas if $\Delta I_{diag} > 0$ $w_{diag} \neq w_{opt}$

Disclaimer
Most of the information of this summary was taken from the lecture slides of Prof. Daniel Kiper and extended with information of the book Wandell, B. (1995) Foundations of Vision. Sunderland, MA: Sinauer. Some little clarifications were taken from searches in the internet. The authors don't claim that this is their intellectual property. It is solely to be used for studying for the UZH/ ETH course "INI 402"

8.3 Decoding a population code

If one looks at the effect of correlation on decoding one sees that decoding is much easier if the visual system has an idea of the amount of correlation between the noises and the stimuli therefore. If the preference of the two neurons is not different so $\Delta I_{diag} = 0$ there is no difference between an uncorrelated and a corrleated response, since the lines of best separation overlap. If however One neuron prefers a certain stimulus ($\Delta I_{diag} > 0$) knowledge about the type of correlation is beneficial, as the lines of best separation do in fact not overlap. These simulations are visible in figure 31.
Another example is when looking at the number of neurons (the population of neurons). The loss of information decreases with increasing population size with a weak correlation, because the neurons can differentiate better and make clearer predictions.

The ideas of correlation are important in the studies of artificial neural network because there we are interested to think about correlations onf neurons and input data in order to improve our predictions using these ANNs.