[SQUEAKING] [RUSTLING] [CLICKING]

JOSH MCDERMOTT: We got sound working. And so I'm going to show you a couple of these movies that didn't quite work last time. This first one is a demonstration of a center-surround receptive field in the retina. So the movie is made in an experiment where there's an electrode that is recording from a retinal ganglion cell. The eye is pointed towards a screen. This is a video of the screen. And onto the screen, they project light. And they will investigate the effect of the pattern of light on the response of the neuron.

So let's see if we can get a sense of what's going on. Oh, and the other thing that I should say is, as is fairly conventional with this kind of thing, the output of the electrode is being fed to an amplifier and a speaker. So you're going to listen to it. So every time there's an action potential, it'll show up as a click that you can hear. So you're going to be listening for the crackling sound of spiking activity.

[VIDEO PLAYBACK]

[CRACKLING]

[END PLAYBACK]

JOSH MCDERMOTT: OK. I'm going to pause real quick. So one thing that you should have noticed is that, as this spot of light is moved around, there's this region of space where, when the spot of light sweeps through there, the response of the neuron increases. But you can also hear that, even when the spot of light is around in other places, there still sometimes are spikes. And that's because there's spontaneous activity that's being emitted by the cell. So now, they've centered the spot of light in the approximate location of the receptive field, the place that's driving the neuron. And now, it's flashed on and off. And you can see that the flashes illicit bursts of spikes. Oops.

[VIDEO PLAYBACK]

[CRACKLING]

JOSH MCDERMOTT: So the point here is that the small circle gives a bigger response than the big circle.

[CRACKLING]

JOSH MCDERMOTT: That's because there's an inhibitory surround.

[CRACKLING]

JOSH MCDERMOTT: The point of this is the dark circle kills the response. And then you get spikes once it disappears.

[CRACKLING]

[END PLAYBACK]

JOSH MCDERMOTT: All right. I think you get the idea. So that's the center-surround receptive field. There was one other that we were going to look at last time that did not work, which I'll show you now, which is a similar kind of experiment, but mapping out an orientation-selective receptive field. So let's check this one out.

[VIDEO PLAYBACK]

[CRACKLING]

JOSH MCDERMOTT: OK, so the region with the X's, the region that increases the response when there's light there. The O's, or the triangles, whatever that is, is the region where the response is suppressed by light.

[CRACKLING]

JOSH MCDERMOTT: You can see it's orientation-selective. So now, you change the orientation and the response changes. And then you hit the right one, you get a bigger response.

[CRACKLING]

[END PLAYBACK]

JOSH MCDERMOTT: So that's how receptive field mapping worked in the old days. So last time, we talked about primary visual cortex. So just to remind everybody, so this is what we're talking about. It's the visual system. Light enters the eye. There's an image formed on the retina. There's a set of layers of cells, the output of which comes from the retinal ganglion cells. That goes through the optic nerve to the LGN, part of the thalamus. And from there, it projects to the visual cortex. And the primary visual cortex is the main projection site of the thalamus. It's the biggest region in the visual cortex.

We talked about how one of the key organizational principles of the visual cortex is retinotopy. So there are maps of the visual field that you can see in the brain. And they distinguish different visual areas, as we'll see later. We talked about the principle of cortical magnification. So one of the consequences for the foveal organization of the retina, where there's a high density of receptors at the center of gaze, is that there's more area in the cortex that's devoted to the center of gaze. And you can see that in various ways.

We then talked about orientation selectivity, including this video that we just saw, and how we can approximate the receptive fields of orientation-selective neurons with linear filters that have these kinds of oriented forms. We then talked about this simple model that's been proposed for how you might create an orientation-selective neuron from the responses of center-surround receptive fields at an earlier stage of the visual system by having appropriate convergence, and then talked about an experiment that provides evidence for this, where the experiment found pairs of neurons in the LGN and V1 that were likely to be connected based on the cross-correlogram, and then looked at the correspondence between the receptive fields and found that the correspondence was pretty good, providing evidence for that simple model.

We then talked about the columnar structure of the visual cortex. Again, a pretty defining characteristic of lots of cortical areas is that they're organized into columns, meaning that, if you are inserting an electrode perpendicular to the cortical surface, the neurons that you'll encounter will tend to have fairly consistent properties.

So in particular, there are orientation columns. So all of the different cells that are sampled on electrode penetration 1 have the same orientation preference. That's what those little line segments are indicating, whereas if you insert the electrode obliquely to the cortical surface, you go through a bunch of columns. And so the orientation preference changes as a function of the electrode position.

We saw some pictures of orientation columns. And then we talked about the ambiguity of the response of a single neuron and how, in this particular example, contrast and orientation are confounded, and how you could resolve that ambiguity using a population code. So if you were able to look across a set of neurons that varied in their preferred orientations, even in a situation where the contrast is varied, such that you could get the same response for all these different orientations in one neuron, if you were able to analyze the population, in particular detect the peak of the distribution, you would be able to infer the correct orientation despite the change in contrast.

And then we ended class by experiencing the tilt aftereffect, where we induced a temporary change in your brain by forcing you or imploring you to stare at the adapting stimulus, which is this oriented grating. And you found that if you stared at this thing for about a minute, and then you looked at the test stimulus, even though the test stimulus is actually vertical, it appears to be a little bit off of vertical. So that's what's known as the tilt aftereffect.

A few details about this-- so you may recall that, when we were doing the adaptation, I instructed you to move your eyes around inside the gray circle. And the reason for this is that adaptation happens at all stages of the visual system. And in particular, it also happens at the retina, even at the level of the photoreceptors.

And so if you maintain very tight fixation on something, you will actually get very strong adaptation of the photoreceptors. And you'll get an afterimage. You'll see the negative of the image that you were looking at. And that might be distracting and might make it a little harder to actually observe the orientation.

But by moving your eyes around in that gray circle, you don't systematically adapt the photoreceptors because you'll be varying the position of your eyes by more than the spatial frequency of the grating. So that's just a trick to get these effects to work pretty well. So we got the tilt aftereffect. And if you would like to, when you go home, you can try adapting to this one. And you will find that you would get the aftereffect in the opposite direction.

All right. So the question is, how can we understand these kinds of effects? What do they tell us about the brain and the visual system? And so in general, we think that aftereffects are due to something that happens in your visual system, whereby there is a decrease in the responsiveness of neurons after a period of prolonged activity.

And so the idea is that, when you are adapting yourself-- so you're staring at this grating. And so the neurons that are responsive to this grating are firing pretty hard for an extended period of time, like a minute. And for some, as of now, unspecified reason, we postulate that they become less responsive as a consequence of having fired pretty intensively for a period of time.

For instance, you might imagine that there's a metabolic resource that's needed for spiking. And that gets temporarily exhausted or depleted a little bit. So we believe that that's really what is underlying this. But how can we understand this in more detail? So here's the idea.

So remember, we just talked about population codes. So the idea of the population code is that you get some stimulus. There's a population of neurons. These triangular-looking things are the tuning curves of a hypothetical set of neurons. This is plotting the response of the neuron as a function of orientation on the x-axis. And then the little line segment at the top is symbolizing the preferred orientation of a given neuron.

So this is the population response normally to the vertical grating. And what you can see here-- so now, the x-axis here, the line segment denotes the preferred orientation of a cell. And you can see that the cell that prefers the vertical orientation gives you the highest response. And then the ones on either side give you a lower response. And you get this bell-shaped curve with a peak at vertical.

And so the hypothesis is that there is a mechanism that analyzes this population response, detects the peak, understands that the peak is occurring at a neuron whose preferred orientation is vertical. And that causes you to perceive a vertical orientation. So that's what we think happens normally.

Now, we did this crazy thing to your visual system, where we made you stare at that thing for a minute. And so the hypothesis is that now the responsiveness of some of the neurons has been altered. And that's symbolized by the tuning curves being lower on the graph. And we hypothesize that the decrease in responsiveness is proportional to how strongly the neurons were firing when you were doing the adaptation. And so in particular, the biggest decrease here is at the orientation that's equal to the adapting orientation. And then there's a graded effect on either side.

So the consequence of that is that, when you subsequently view the vertical grating again, the population response has now been altered. So normally, the peak of the response occurs in the vertical neuron. But because of the adaptation, you can see that's no longer the case.

So we adapted the stuff over here. And that means that the distribution has shifted. And now, the peak of the distribution is this orientation here, which is off of vertical in the opposite direction. And so if you believe that there is some mechanism that can detect the peak and use that to infer the orientation, then that predicts that the perceived orientation will be shifted in the direction that you experience after you've adapted. So that's typically how we explain aftereffects and how they are believed to provide evidence for these population codes. Yeah?

STUDENT: So in the last slide, why are all of the peaks at the same height?

JOSH MCDERMOTT: So these are tuning curves. So that's just telling you how much the cell would respond to each of these orientations. This is the actual response of the different cells to that one orientation. But the point is that all of these cells initially are equally responsive. But then they become differentially responsive temporarily following adaptation. Any other questions about this?

So there's lots of aftereffects, especially in vision. They're actually like a lot more powerful in vision than in audition. It's interesting to think about why that might be the case. But we'll encounter lots of other aftereffects. And they were classically a pretty common tool that was used by perception scientists to try to infer things that were happening in the brain. So the tilt aftereffect is one. We'll also experience the motion aftereffect when we start to talk about motion. That one's really powerful, lots of aftereffects.

So we've got we've got these orientation-selective neurons. We can adapt them. And we discussed how they have these receptive fields that you can summarize with pictures like this. And the fact that we can draw a picture like this really is a reflection of the fact that the neuron is well-approximated by a linear filter. So a linear filter has this property that you can define coefficients at different positions in space. And that does a good job of predicting the response.

And so it's pretty common to observe responses that are-- receptive fields that are either even or odd. And you might wonder, OK, well, why do we have these different types of receptive fields? And so in the early days, the people who initially discovered orientation selectivity, Hubel and Wiesel, tended to interpret these receptive fields with Intuitive verbal descriptions.

So they would talk about these kinds of receptive fields as edge detectors and these kinds of receptive fields as bar detectors. And so there's this very appealing intuitive theory of vision that of fit into, where the idea is that you're going to initially detect edges, and then use those edges to detect and recognize objects. And so initially, when this was discovered, these orientation-selective receptive fields fit into that view.

So the challenge with that way of thinking about things-- well, there's lots of challenges. But one of the challenges is that edges are, in practice, very hard to detect with local operators. So there was lots of work in the early days of computer vision-- 1970s, 1980s-- trying to figure out how to detect edges. And they would include ingredients like these linear filters, and then do some other stuff because the raw responses of the linear filters usually wasn't enough. And they still didn't work great.

So this is an example of a very well-known classical edge detector known as the Canny edge detector, run on this particular image. And so you can see that you get things out that are not ridiculous. But the output really is pretty far from maybe providing what you might think would be a complete description of all of the edges that you can see in the image.

So there's some that get completely missed, like the edge of the chin. And so in general, edge detection is a fairly-- is a pretty hard problem. And I think one very likely possibility is that this intuitive idea that one could separate the stages of detecting edges and detecting objects is not a great idea, that, in practice, like the edges are inferred along with the objects. But that's a longer discussion we can have later.

So that was these original intuitions about orientation selectivity. It's got some issues. Another challenge to that way of thinking about things is that if you actually look at the responses of the kinds of filters that you see in the visual system in the way simple cells, if you look at their responses to actual edges and actual lines, the responses are pretty complicated.

And so here's an example, where we've got an input image. So that's an image. It's actually a one-dimensional slice through the image. So the y-axis would be intensity. And the x-axis would just be one spatial dimension. And so you can see that there are two step edges where the intensity goes up, and then two lines-- one light and one dark.

And then B and C plot the convolution of an odd symmetric filter-- that's B-- and an even symmetric filter-- that's C. So we're measuring the response of the filter at every position in the image. And that's what a convolution does. And again, naively, you might expect that an edge detector is just going to give you a delta function at the edges. And the line detector would give you a delta function at the lines.

But in actuality, all of the filters kind of respond at all of the features. But the responses are complicated and wiggly. So it's not to say that you might be able to take these, and then do something else, and maybe infer something. But at any rate, the responses don't completely, straightforwardly detect edges and lines in an intuitive way.

And so I would say, so there have been multiple iterations of attempts to explain and understand orientation selectivity in computational terms, some of which are beyond the scope of the class. I would say the relation of simple cells and orientation selectivity to function remains challenging to verbalize, to explain in an intuitive way.

But it's nonetheless the case that similar looking filters pretty consistently emerge in the initial stages of systems that are trained on vision tasks. So this is an example. So these are the first layer filters of a convolutional neural network optimized to recognize objects. This is a very famous convolutional neural network. This was the first one that produced really impressive performance on ImageNet, a classical object recognition task.

This was back in 2012. And this opened the gates of deep learning and got everybody really excited about deep learning. Each one of these is a different filter. And you can see that they look a lot like simple cells, in the sense that they're orientation-selective to tune to different orientations. And so this is a very common observation. So you very often see orientation selectivity emerging as the initial stage of analysis in systems that are trained on sensory problems, in particular vision problems.

So mathematically, if we actually want to summarize the form of a simple cell, it's common to describe them as Gabor functions. So a Gabor function is the product of a sinusoid with a Gaussian. So we've got a sinusoid up here and a Gaussian. And we just pointwise multiply them. And you get this thing, where it starts out at 0. And then the wiggles increase at some rate. And then they decrease again. This is a Gabor function displayed as an actual image.

So both of these types of receptive fields, even and odd, can be described as Gabor functions. The only difference is whether you use a sine or cosine, because the even and odd functions are 90 degrees out of phase. So that's at least a mathematical description of what these simple cells are doing.

Now, the other wrinkle to the story is that, in addition to simple cells, in primary visual cortex, there's another type of cell that became known as complex cells. And complex cells, they differ from simple cells in that they respond to-- so they're still orientation-selective, but they respond to an oriented stimulus no matter where it falls in the receptive field.

So remember in that video that we just saw, where they were mapping out the orientation-selective receptive field, it was very sensitive to where that bar was placed. So if you got it in exactly the right place, you'd get a big response. But then if you shifted it a little bit over, the response would decrease. So that's characteristic of a simple cell and consistent with the idea that there's a linear filter that's got this excitatory region. And so if you hit that with light, you get a big response. And if you're in the inhibitory side lobe, you get a small response.

But complex cells behave very differently. They respond no matter where the stimulus falls on the receptive field. And as a consequence, they can't be decomposed into these excitatory and inhibitory response zones. So here's just an example picture that differentiates the two types of cells.

So we've got the receptive field area in gray. The stimulus is the orange bar. And the simple cell response, you can see, is high when the orange bar is at the right orientation and in exactly the right position. And then if you shift the orange bar a little bit to the right, the simple cell response drops off.

Now, it's also orientation-selective. So if you get the wrong orientation, you also get a lower response. The complex cell response, by contrast, is high at both of the positions. But it's still orientation-selective. So this is, again, an empirical discovery in primary visual cortex. You've got these two types of neurons.

And it turns out that complex cell responses can be modeled by combining even and odd simple cell responses. And when you model simple cells as Gabor functions, this turns out to be mathematically very simple. So an even receptive field would be a Gabor that has a cosine in it. So G here is a Gaussian. And that's a cosine with some spatial frequency. That's the odd function with sine.

And so if you just square the responses of the even and odd receptive fields and then add them together, because of the trigonometric identity that sine squared plus cosine squared equals 1, you end up getting this very simple expression, where the sinusoids drop out of the expression. And you're just left with a Gaussian spatial dependency.

All right. So this is what is often known as the energy model of complex cells. And we now have a second-- or of a third row to our primary visual cortex response graph here for edges and lines. And the bottom row is energy computed from the combination of the odd symmetric and the even symmetric responses by squaring them and summing them. And now, you get this nice little bump anytime there's an image feature.

So this energy model, it provides a reasonable model of a complex cell. So there aren't any more of these excitatory and inhibitory lobes. It inherits the tuning that is endowed from simple cell input. And so I should say that what's kind of implicit in here, which I should make explicit, is the idea that the complex cell is created by combining the responses of simple cells.

So just as the simple cell, we believe, could be created by wiring together center-surround receptive fields in the LGN, the idea is that the complex cell is created by wiring together simple cells that have the same orientation selectivity and different phase, or even and odd receptive fields. So that's the notion. And the consequence is that they kind of inherit some of the tuning properties-- orientation selectivity. Sometimes, there's also motion selectivity. But the response becomes nonlinear. Any questions about simple versus complex cells?

So this is a picture that captures like a fairly standard model of early vision. When people talk about early vision, normally they mean roughly up to primary visual cortex. So it's capturing early vision using convolution. So we've got an image there. It's an image of beans. And then that image is convolved at each step with an operator that is embodying a particular stage of visual processing.

So we've got center-surround cells here that capture what happens in the retina and the LGN, where the center-surround receptive fields are different sizes. Then there's orientation-selective filtering, and then the energy extraction step. So again, the idea is that the convolution operation is used to represent the response of a large population of neurons that have the same kind of receptive field, but positioned at different spatial locations.

And so you can think of the response of a stage of the visual system kind of as a function of an image. So you have the image as an input. And you get this response out that looks like an image. And the response is high in these different places where the original image, in this case, had orientation energy that was at the right orientation for that cell, and same here, and same here. And if you look back in the image, you can probably convince yourself that's the case.

But then when you get to the energy quantity, you can see that, well, the energy is high any place that either has high or very low activity in the simple cell responses, because you get the squaring operation and summing. And so that causes it to be spatially a little bit more coarse.

Now, you might wonder, OK, what good is this? What does this have to do with vision? That's a legitimate question. And so I think the broader comment that I would make here is that there's a bunch of different angles at which to try to understand perception. And one of those angles is to go into a sensory system, and to look at what's there, and to try to measure what's there, and characterize that. And that's what we've been talking about over the past couple of lectures.

And that actually had some success in the first few stages of the visual system, in the sense that the measurements often-- they ended up being mathematically interpretable. And people were able to come up with mathematical descriptions of the responses of the neurons that were pretty good, in the sense of being able to predict the responses of the neurons.

Now, does that tell us how these things help us see? No, they don't. These are just descriptions of parts of the system. But they are made with the hope and expectation that they will end up being relevant to our understanding of how we see. Now, the other way to get at perception is to start with perception itself, to think about the things that we can do, and to measure that, and to try to come up with theories of that, and then eventually try to connect that back to neurons.

So remember, we talked about different levels of understanding. There's the level of the computation, the algorithm, and the implementation. At various points in the history of this field, sometimes people start with the implementation. So they're actually measuring from neurons, trying to document that. And some people really believe that that's where you start. And that's the best place to get rolling.

In other cases, people actually start by trying to understand the computation, trying to understand the function that the system is carrying out and how to understand that in mathematical terms, and then worrying about the implementation later. And different sections of this course will have more of one perspective and more of the other. And ideally, all of these things are going to come together. And they don't always, at least not yet. But hopefully, eventually, that will be the case.

So big ideas so far-- we've got these stages of linear filtering in the retina and in primary visual cortex with simple cells, and then this one kind of non-linear operation that happens in complex cells that we can capture with the notion of energy. Any questions about that before I move on? Yeah?

STUDENT: Can you repeat again what energy expresses in relation to the original image, with the beans example?

JOSH MCDERMOTT: Yeah, so the energy will be high whenever there is-- whenever an oriented filter that's at the orientation, which you're measuring the energy, whenever an oriented filter, whether it's even or odd, provides a large response. Yeah, so I think, here, see how there's a white blob here? And it's because there's a oriented bean that's in that place in the image.

So there's a place where the edge-- unfortunately, I need a point here. But there's a place where it goes from dark to light, and then another place where it goes from light to dark. And so those will produce-- one of those would produce a positive response in an odd receptive field. The other would produce a negative response in the odd receptive field. But the energy squares that and sums that. So it doesn't really care about the sign. It's just telling you there's stuff there that's at that orientation. So that's one way to maybe intuitively describe it. Yeah?

STUDENT: So I think it's true that there are center-around neurons in the LGN that detect color. How does that get incorporated in the orientation and energy extraction steps?

JOSH MCDERMOTT: Well, in this model, there's no color. This is a black and white model. In the visual system, I mean, we're going to talk a little bit more about that when we get to color. But that's part of the story. And in fact, I mean, one of the things that has been debated is the extent to which orientation-selective mechanisms are also color-selective, or whether color vision is maybe a slightly different thing.

But we'll see evidence of that. I think we now mostly know that you can find color selectivity in conjunction with orientation selectivity. Yeah, there are some interesting differences, like the spatial resolution of color vision tends to be pretty poor. So there are indications that there is some functional separation between certain kinds of high resolution, pattern analysis, and color. But more to come on that.

So you get to the visual cortex. And you see some new stuff that was not present in the responses of the retina or the thalamus. The first example of that we've been talking about is orientation selectivity. Another example is that you find neurons that respond to both eyes.

So prior to primary visual cortex, all of the cells get input from either the right eye or the left eye. So of course, in the retina, you're either in the right eye or the left eye. But then you get to the LGN and different layers of the LGN are carrying information from either the right or the left eye. But then in primary visual cortex, you see neurons that get input from both eyes and that will respond more when both eyes are stimulated. And those are called binocular neurons.

But it's nonetheless the case that many cells are still driven more by one eye than the other. And so this is a histogram in two different species that plots the extent to which there is ocular dominance. So again, I don't know why this is so blurry. But on this end would indicate that the neuron is dominated by the contralateral eye-- contralateral is on opposite side. On this end, it would mean that it's dominated by the ipsilateral eye.

And the point is that there's a lot of mass out here and out here. So a lot of neurons are dominated by one eye. And that shows up in another structural feature of the visual cortex, which is ocular dominance columns. And so that is that, again, if you're looking down at the cortex-- this is the cortical surface. And this is the result of an experiment where people were able to image the activity in the visual cortex while an animal had one eye open.

So one eye is stimulated and the other eye is closed. And you see this patchy pattern of activation, where there are bits of the cortex that are a lot more active than others. That's ocular dominance. So the response in the white bits of the cortex is driven by the eye that's open predominantly. So we've got orientation columns and ocular dominance columns. And they're interspersed in some kind of way.

So this is from optical imaging in animals. You can also see this stuff in humans. This is an fMRI experiment that measured ocular dominance columns. This is two particular subjects, and then overlaid them on orientation columns. So these columns gave rise to this cartoon model of the visual cortex, where the idea is that you could think of the visual cortex as a set of hypercolumns, where each hypercolumn contains all of the different kinds of measurements that you would want to make at one point in space. And those measurements are different orientations, as you can see here. And also, the left and the right eye. And this is also sometimes referred to as the ice cube model.

So how does all of this hooked up to the LGN? So remember, this is the LGN. The LGN's got these six layers. The top four are the parvocellular layers. The bottom two are the magnocellular layers. This is a picture, though, that shows the labeling from one eye. And so you can see that alternate layers are getting input from one eye. So each layer gets input from either the left eye or the right eye. And that one, that one, that one are getting it from one eye, not sure which one.

All right. So this is showing how the LGN provides input to V1. So in general, most cortical areas have the input coming in at layer 4. So layer 4 is the input layer for the cortex, typically. And in V1, a lot of the input comes into layer 4C. And so we've got these two layers of the LGN-- one for the right eye one for the left eye. And then those have these alternating spatial projections that correspond to ocular dominance columns.

The one other wrinkle is that-- so remember how we talked about how we've got parvocellular layers and magnocellular layers. Remember, we talked about these lesion studies where, if you selectively lesion the parvocellular layers, you get deficits in color perception. If you selectively lesion the magnocellular layers, you get deficits in motion perception.

And so, unsurprisingly, the responses from the parvocellular and the magnocellular layers remain segregated. And they, in fact, project to different sublayers of this layer 4C. So 4C alpha is where the magnocellular layers end up. And 4C beta is where the parvocellular layers end up.

So in general, the central theme here is, especially in the early stages of the visual system, there's really fairly striking segregation of function. You've got these distinct cell classes. There's a small number of cell classes in the retina. They have different properties. Those project to different parts of the ln that have different properties, different impact on behavior. And then that remains segregated through V1 and also a little bit beyond.

So a lot of the input comes in here at layer 4C. And then this is just a diagram. I don't expect people to memorize this. This is just included for cultural literacy. You can see connections between different layers of V1. So you get the input coming in here at layer 4, and then projections to other layers. And then you can see there are certain layers that are like the output layers that tend to send responses to other cortical areas. There's other layers that tend to send responses to non-cortical brain structures. So there's just a lot of interesting anatomical segregation.

And every decade or so, more is learned about this stuff. And so the pictures get more complicated. So this was the diagram that you would find in a textbook back in 1995, where pretty much people just talked about magnocellular and parvocellular layers from the LGN.

This is a more recent diagram. And so now, there are these koniocellular cells that are in between the parvocellular layers and the magnocellular layers. And those project to a different place. So as time goes on, these things will probably get more and more complicated.

Another thing that you see in the cortex is direction selectivity. So here's an example of a cell that is selective for the direction of motion. So the dashed rectangle here is the approximate location of the receptive field, the part of space that causes the neuron to respond. There's a bar that's being moved back and forth. You can see that there is a preferred orientation, which is this oblique orientation. But you can also see that the response is much bigger when it moves in one direction than when it moves in another direction. So direction selectivity is pretty common and, again, present in layers that tend to get input from magnocellular parts of the LGN.

And there's other complicated stuff. This is a property that is often called endstopping. So if you have an orientation-selective neuron that's, in this case, selective or horizontal, if you make the line too long, the response actually decreases. So there's an inhibitory surround on top of the orientation-selective response. So that's also a pretty common thing that you see in the cortex, is inhibitory surrounds, even though it's not just a center-surround receptive field. You have some kind of complicated tuning, and then a surround on top of that.

So the key take-home themes that you should take away from this lecture are the idea of retinotopy as an organizational principle for the visual system and cortical magnification-- so again, in the retina, we've got the foveal organization of the receptor lattice. And that translates into cortical magnification when you get to the cortex-- orientation selectivity and the idea of how you would create orientation selectivity in a neural network with input from center surround receptive fields, simple versus complex cells-- so simple cells are approximately linear.

The receptive field consists of excitatory and inhibitory regions that are spatially localized. Complex cells are nonlinear. They don't have these distinct spatial subregions-- the idea of population codes and how they are used to infer aftereffects and how after effects provide evidence for population codes-- the idea that if you analyze a population of neurons, you can estimate stimulus properties-- the use of convolution to simulate a population of neurons that has the same kind of receptive field at different spatial positions, columnar structure as a central organizational principle of the brain, these anatomical regularities that we've been talking about, about how the connections are very non-random, and then the zoo of response properties that emerge in the cortex, where there's a whole bunch of stuff that you see in primary visual cortex that's not evident or much less evident in the retina and the thalamus.

So we talked about orientation selectivity, binocularity, direction selectivity, endstopping, stuff like that. Any questions on primary visual cortex?

So we've just been talking about how the receptive fields that you see in V1, they vary in a bunch of different respects-- orientation tuning, phase, even versus odd, endstopping, binocularity, direction selectivity, whether it's simple or complex. And they also vary in what's called spatial scale.

So this is a picture that shows receptive fields measured at a given point in cortex that represents a particular position. So each one of these boxes represents a receptive field. So it's roughly the region of space that drives the response of the neuron. And so the point is that the boxes are different sizes. So some of them are kind of small. And some of them are kind of big. So in any given region of the cortex, the receptive fields will all be at pretty much the same place. But there's a fair bit of variation in their size.

So what does this mean? Well, remember back when we were talking about sound and hearing? We talked about Fourier's theorem and the fact that Fourier's theorem shows that any signal can be written as a sum of sine waves of different frequencies and phases. And so we were talking about this in the context of sound, where this dimension was time and this dimension was pressure. But Fourier's theorem is just a general fact about any kind of signal.

So an image is just an array of numbers. So it's a signal. Fourier's theorem applies just as it does to sound. And so it follows that any image can be decomposed into a sum of sine wave gratings of different orientations, frequencies, and phases. And the different spatial frequencies capture what are called different scales of image structure.

So what do we mean by that? Well, this is a sinusoidal component of an image. It's a sinusoidal grating. So what that means is that the image intensity varies sinusoidally, in this case across the horizontal dimension. So back when we were talking about sound, I was telling you that the y-axis was pressure and the x-axis was time. Now that we're talking about images, the y-axis will be intensity and the x-axis would be space or position in the image.

So here's one sine wave grating. These are gratings that have different spatial frequencies. So you can see that the rate at which the intensity varies across space is different in the three cases. So spatial frequency is typically measured in cycles per degree. And so this would be a low spatial frequency grating. And this would be a higher spatial frequency grating.

The grating can also vary in phase. So these gratings have the same frequency, but different phases. So the exact position of the ripples is different in the three cases. And so you can take these sine wave grating images and you can add them. So here's one grating. And here's another grating that's three times the spatial frequency of the first one. And you add them up. And you get this more complicated thing.

So here's what happens when you add up these two gratings in different phase relationships. So these are the same two gratings, but with different phases. And so here's the image that results in these two cases. And then here's a slice through the image. So now, again, the y-axis is intensity and the x-axis is space. And so it's a little easier to inspect. And so you get these different shapes that result from different phases.

All right. Now, one of the things that we talked a lot about with sound and that you got some experience with on the problem sets and possibly in your illusion labs is filtering. So filtering is an operation that changes the frequency content of a signal. And so images can also be filtered.

Now, there's one wrinkle of doing this in the image domain that is sometimes a little bit hard to think about. And that is that images are two dimensional. And so there's actually two dimensions of spatial frequency, x and y, whereas with sound, when we plotted the amplitude spectrum or the power spectrum, we'd get a one-dimensional thing. So you have power on the y-axis and frequency on the x-axis.

So now, with images, there's two dimensions of frequency. And so you have to display the power spectrum as a three-dimensional plot. So now, this is amplitude here. And this is one dimension of spatial frequency. And this is another dimension of spatial frequency. And so what this is showing is the frequency representation of this image of a face.

And what this shows-- so at the center, that corresponds to the 0 spatial frequency, or the DC component. And then the frequencies around that are the low spatial frequencies. And then as you get out into the extremes here, those are higher spatial frequencies. And so in general, natural images tend to have frequency spectra that are lowpass. So they tend to have more power at low spatial frequencies and less at high. And so this is just another example of that.

So that's the frequency representation of the image. And just as if you have an audio waveform, you can get the power spectrum by taking the fast Fourier transform. There's a Fourier transform that can be performed on an image. It's just a two-dimensional Fourier transform

And so you can have filters. And these are transfer functions of two filters. So this is a lowpass filter. So what this does is it passes all of the stuff that's right in the center of the frequency plane and attenuates everything beyond that. This is a highpass filter that passes just the very high spatial frequencies and kills off the lowest spatial frequencies.

And so we can take a lowpass filter, apply it to the image, and this is what you get. You get something that is blurry. This is a highpass filter applied to the image. And you get something that looks like this, where a lot of the edges are accentuated. And so intuitively, what has happened here is that the lowpass filter is preserving the low spatial frequencies. So those are like the parts of the image that vary slowly across space. And the highpass filter is eliminating those and just preserving the parts that vary very rapidly as a function of space.

So images can be filtered. So here are some examples. Here's another face image with the highpass filtered version and the low passfiltered version. Again, the lowpass filtered versions just look blurry. Here's another image with the low frequency component and the high frequency component.

And this is an image that is built up from a lowpass component, a bandpass component-- that's number 1-- and a high pass component-- that's number 2. And you get something that looks more like a full resolution image. Here's another example of a photograph of flowers that's been decomposed into a lowpass, bandpass, and highpass components.

So a lot of these same concepts that we talked about when we were talking about sound can also be applied to images. But now, we're dealing with spatial frequency instead of audio frequency, where things are varying over time. And it's two-dimensional, which always makes everything more confusing.

All right. So this is math essentially. So we've got signals. The Fourier transform lets us decompose them into frequencies. We can filter them. So what's the relevance for vision? So the natural question is, does the visual system decompose images into spatial frequency components? So we saw that the cochlea decomposes sound, to some extent, into audio frequency components. Maybe the visual system does something similar.

So to look at this, we are going to get into the business of measuring contrast detection thresholds. So for that, we have to define contrast. So intuitively, contrast is the difference in intensity across some points in space. So if we have a sinusoidal grating where the luminance is varying spatially, sinusoidally, there will be a minimum luminance and a maximum luminance. And the contrast of the grating is defined by the difference between Lmax and Lmin divided by the sum. And so, by definition, the upper limit of the contrast can be 1.

So that's what we're going to define. And what we want to measure is your contrast detection threshold. So how much contrast is needed for you to be able to detect that there is a grating there-- that is, that there's variation in intensity? And so it's kind of obvious that, when the contrast is high, it's going to be easy. As the contrast gets lower, it'll start to get harder and harder to tell that there are ripples in the image. And at some point, it will become impossible. So we'll measure the threshold using the psychophysical methods that you mastered a few lectures ago. Yeah?

STUDENT: Does the number of bands matter in detecting contrast?

JOSH MCDERMOTT: That is one of the questions we are about to answer. Yes, it's a great question. So we're going to do an experiment where we will ask, what is the smallest amount of contrast that is visible? So these are gratings that vary in their contrast. And so we measure the threshold. And then typically, in this particular subfield, we deal with contrast sensitivity rather than contrast threshold. I apologize for the typo here. So sensitivity is just the reciprocal of threshold. So in general, if you're more sensitive, that means your threshold is lower.

So the question was posed, does the threshold depend on spatial frequency? And so this is an image that answers that question. So this is an image that varies in spatial frequency from very low on the left to very high on the right. And it varies in contrast, going from high contrast at the bottom up to low contrast at the top.

And what you should be able to see by looking at this is that your contrast sensitivity is not uniform across spatial frequency. Rather, it's best for medium spatial frequencies and is worse when the spatial frequencies get too low or get too high. Everybody see of inverted U shape? All right.

So that is what is known as the contrast sensitivity function. So the contrast sensitivity function is a function that plots contrast sensitivity as a function of spatial frequency. So remember, sensitivity is 1 over the threshold. So when the number is high, that means your threshold is low. And you can see that your sensitivity is highest at these medium spatial frequencies, like three or four cycles per degree, and then gets lower if the spatial frequency gets lower and higher if it gets higher. And that corresponds to what you saw in that image that we just demoed.

And so the consequence of that is that you can see the image variation in this region. And even though there is physical luminance variation up here, that is invisible to you. So this is just a property of your visual system. But what about our original question of whether the visual system decomposes images into different spatial frequency components? Are there frequency channels, kind of like we see in the auditory system?

All right. And so to get at this, we're going to use a classic tool in the perceptual scientist toolbox, which is adaptation. And so what we're going to do is adapt ourselves to a single spatial frequency. So we're going to stare at a sine wave grating for about a minute. That's usually the magic number. And there are naively two hypotheses.

One hypothesis is that the visual system doesn't have these spatial frequency channels. And that would predict that, if you're adapting to a grating, you might just get a general decrease in sensitivity. So it might get harder to see contrast irrespective of the spatial frequency. The other possibility is that, if your visual system does contain these different channels that are specific to particular ranges of spatial frequencies, that if you adapt to one, you will decrease the responsiveness in that one channel, kind of like what we think happens when you adapt to an orientation that causes the tilt aftereffect.

And that would have the consequence then-- so if we adapt ourselves and then measure contrast sensitivity, we should see a localized decrease in contrast sensitivity at the adapting frequency rather than a global decrease. So let's try it. So what I want you to do-- so first, just glance at B. So the image on B is the one that we're going to use to measure your contrast sensitivity function.

Now, let me get out my stopwatch. I want you to adapt your visual system by staring at the red circle. And so again, look around in the red circle, because we don't want to adapt our photoreceptors. We're trying to adapt spatial frequency channels. So just look around at the red circle. You can think about whatever you want. Just keep your eyes in the red circle.

And eventually, we're going to then look over at that other image. And the two possibilities that we're going to try to evaluate is whether the overall the profile of your sensitivity shifts down. Again, don't look now. Just keep looking at the red circle, OK? We're going to ask ourselves whether the overall sensitivity drops or whether there's a little bite taken out of the contrast sensitivity function around the adaptive spatial frequency.

All right. We're going to do this for about 15 more seconds. Hang in there. Just keep looking at that red circle. 5, 4, 3, 2, 1. All right. Look over at your contrast sensitivity function.

And what you should notice is that there's now this little bite taken out of the area around the adaptive frequency. Is that what everybody sees? Raise your hand if that's what you see. Most of you. Not everybody. Yeah, OK. So we may not have adapted long enough for everyone. But that's the typically experienced result, that you should have seen a profile of visibility that looks more like this, rather than the upside down U shape that you typically see.

So this is the result. And the way that you would do this experiment in the lab is-- it's a bit of a tedious experiment because you've got to measure detection thresholds at all these different spatial frequencies. And you'll want to measure that under normal conditions, and then under conditions where you adapt people.

And so to measure this with adaptation, what you'd have to do is force people to look at the adapting stimulus, and then periodically make them do a detection trial, adapt them some more, have them do some more detection trials. You have to keep them adapted in order to do these measurements. So it's a long experiment. But the outcome of that would be a contrast sensitivity function that looks like this, where sensitivity is reduced in a local region of the spatial frequency spectrum.

So this suggests that this thing that we measure, that's called the contrast sensitivity function, represents the envelope of a whole bunch of different channels rather than a single mechanism, a single receptive field. Now, this doesn't tell us where these channels might originate in the visual system. And that's a question that we can ask.

So one possibility is that these spatial frequency channels are actually due to center-surround receptive fields, like potentially in the retina or the LGN. And so the idea here is that there's a center-surround receptive field, shown here. And it's like a good match to a particular spatial frequency, where the center of the receptive field aligns with the peak of a sine wave grating. And then the surround aligns with the troughs, whereas if the spatial frequency is lower or higher, it's a poor match to the receptive field. And so you get a lower response. So this is just the same linear filter responses that we've seen.

So the point is that you could potentially explain this with these center-surround receptive fields. So does anybody have an idea for how we could test that theory of spatial frequency channels? How could how could we test whether your spatial frequency selectivity is due to center-surround receptive fields?

OK, I have an idea, actually. What if we adapt to a horizontal orientation of the same spatial frequency, and then measure contrast sensitivity with a vertical orientation? So if the adaptation is occurring in a center-surround receptive field, then, if you adapt to this, it should transfer to this.

Let's try it. All right. So we're going to adapt. Stare at this. So the procedure here is we're going to stare at this red circle for a minute. So start staring. And then we're going to look up again and see if it looks any different. It's very important to have your eyes open for this. If your eyes are closed, it will not work. So stare at the red circle down here. Again, move your eyes around within the red circle to avoid getting a big afterimage. OK, we've got another 30 seconds to go.

All right, 20 more seconds. Keep staring. All right, 10 more seconds. 8, 7, 6, 5. 4, 3, 2 1. Now, measure your contrast sensitivity function. how many people see the bite taken out of it now? Nobody? OK. So what we conclude from this is that the effect that we previously witnessed is dependent on some kind of orientation-tuned mechanism.

So it's not occurring at the site of center-surround receptive fields. So it's probably not in the retina or the LGN. So this suggests that the adaptation effects are probably mediated cortically because that's where orientation selectivity arises. And so the idea here is that these channels correspond to neurons or groups of neurons that have similar properties.

So we just measured all this stuff with sine wave gratings. But you can also find evidence for spatial frequency-selective channels with masking. So that's how we really investigate frequency selectivity in hearing. Remember, we talked a lot about masking, and their ability to detect a tone in noise, and how that depended on the bandwidth of the noise, and stuff like that. So this is the same idea, but just applied division.

And so what you're seeing here are letters that are superimposed on noise that has been filtered into different spatial frequency ranges, from very low spatial frequencies up at the top to high spatial frequencies down at the bottom. And so the letters are exactly the same. The contrast of the noise is the same across all of the examples. All that varies is the spatial frequency content of the noise.

And what you are supposed to be able to see is that it's pretty easy to read the letters at the bottom and also at the very top. But then there's this range in the middle where it gets hard to detect the letters. So this is evidence that-- again, more evidence that there is a spatial frequency-selective channel that limits, in this case, your ability to detect the letters.

And you can also see this result in a similar kind of display that lets you see the contrast sensitivity function, like structure. So now, we have these different rows of letters that are decreasing in contrast. And the different columns here correspond to different spatial frequencies of noise. And so the point is that your ability to see the letters extends further down on the far left and the far right, when you're dealing with low spatial frequency noise or high spatial frequency noise, and is impaired most by these middle spatial frequencies. All right, any questions about that?

So some nomenclature in terms of talking about spatial frequency channels-- I mean, a lot of this will be fairly familiar, just from the fact that we covered analogous concepts when we talked about audition. But filters are usually characterized in terms of their bandwidth. So usually, the bandwidth is defined as the ratio of the frequencies at which the maximum contrast sensitivity is obtained-- sorry, at which half the maximum contrast sensitivity is attained. Usually, it's measured in octaves.

So remember, an octave is a doubling in frequency-- so 1 to 2 to 4. So here's two examples of spatial frequency channels. And you can see this one here has got peak sensitivity at 2.3 cycles per degree and a bandwidth of about 2 octaves. This one's got a bandwidth of 1.3 octaves.

In general, the spatial frequency channels that you see in the visual system are wider than the filters of the cochlea. So the cochlea, roughly speaking, filters are maybe a third of an octave, depending on where you are on the cochlea. And the spatial frequency channels in vision are typically in excess of an octave. So the frequency resolution is not nearly as tight. But it's nonetheless there.

So we just did that experiment that suggests that the frequency channels are cortically mediated. And there's now quite a lot of neurophysiology that supports that. So Rus de Valois was the neurophysiologist who did a lot of work in this particular area. And these are contrast sensitivity functions of individual neurons, so single neurons in V1. And you can see that the best spatial frequency of the different neurons varies depending on which neuron you're recording from. Of course, the overall contrast sensitivity also varies. But you can think of each neuron as mediating a spatial frequency channel.

So contrast sensitivity functions have been measured in lots of people at lots of ages. Contrast sensitivity improves pretty dramatically over the first few months of life, especially for high spatial frequencies. So this is the contrast sensitivity function displayed for an adult at the top, and then for different ages of life, starting at four weeks.

And so you can see that, in general, sensitivity improves. But in particular, at these high spatial frequencies, you can see that a month-old baby just can't really detect this kind of stuff at all. And as you mature, you become able to see these higher spatial frequencies. So in general, baby vision is fairly blurry.

All right. Now, so the idea of these spatial frequency channels is that you get this image that comes into the eye. And then it gets decomposed, to some extent, into these frequency components. However, there's lots of reasons to think that those different spatial frequency channels don't remain independent all the way up to object recognition. And this is a classic illusion that provides a demonstration of this.

So how many people know who this is? A few. Yeah, OK. How many people have seen this before? OK, no one's seen this before. All right. So this is a picture of a famous person. But it's very pixelated. And that makes it hard to recognize. But if you squint your eyes, that has the effect of blurring the image. And the effect here that you're supposed to note is that, if you squint your eyes or blur the image, which is essentially providing a lowpass filter-- so you're removing the high spatial frequency components-- the image becomes more recognizable.

And so the inference that we make from this-- so what's happened here, when the image is pixelated, is that there are high spatial frequency components that are introduced that are incorrect. Those are not the high spatial frequency components that were in this original painting or photograph of Abraham Lincoln. They're just artifacts of the pixelation.

And when you squint your eyes, you are filtering out some of those high spatial frequency components. And the fact that that makes the image more recognizable is evidence that, although we think early in the visual system, the high and the low spatial frequencies are decomposed and representing these different channels, they then must interact in order to mediate object recognition.

And in this case, the incorrect high spatial frequencies are interfering with your ability to extract information from the low spatial frequencies. All right. So evidence that these different spatial scales don't remain independent all the way up through the visual system. Any questions about that?

All right. So just to summarize what we talked about today, key ideas are that images can be decomposed into spatial frequencies. That's just a mathematical fact. Evidence that there are multiple spatial frequency channels comes from experiments on contrast adaptations. So if you adapt to a sine wave grating, this results in focal sensitivity decreases. So specifically, you become less sensitive to spatial frequencies around that of the adapter rather than just generally less sensitive. That adaptation is orientation-specific, which suggests it probably happens cortically.

We think these spatial frequency channels are useful for encoding different spatial scales in images. And so the neurons that we see in V1 are tuned for both orientation-- that's what we talked about in the previous lecture-- and spatial frequency. And then we talked more broadly about the contrast sensitivity function as a thing that you can measure, that is what is used to actually see these effects of adaptation.

All right. That's all I've got for you.