NANCY KANWISHER: Auditory cortex is fun to study, because very few people do it. So you study vision, you have to read hundreds of papers before you get off the ground. You study audition, you can read three papers, then you get to play. It's great. So there's consensus about tonotopy, I mentioned this before. This is an inflated brain, oriented like this top of the temporal lobe. High, low, high frequencies. OK. That's like retinotopy, but for primary auditory cortex.

So this has been known forever. Many people have reported it in animals and humans. Oops, we'll skip the high and low sounds, right. So there are lots of claims about the organization of the rest of auditory cortex outside that. But basically, there's no consensus. You know, nobody knows how it's organized.

So what we set out to do, was to take a very different approach from everything I've talked about so far. We said, let's kind of try to figure out how high level auditory cortex is organized. Not by coming up with one fancy little hypothesis, and a beautifully designed pair of contrast conditions to test each little hypothesis. What we usually do. Let's just scan people listening to lots of stuff, and use some data driven method to kind of shake the data. And see what falls out. OK. To be really technical about it.

OK. So when I say we, this is really Sam Norman-Haignere did all of this. And as this project got more and more mathematically sophisticated, I got more and more taken over by Josh McDermott, who knows about audition. And knows a lot of fancy math, much more than I do. They're fabulous collaborators. OK. So basically, what do we do? The first thing you want to do, especially when you're using these data driven methods to broadly characterize a region of the brain, is a major source of bias in the structure you discover, is the stimuli you use.

So with vision, you have a real problem. You can't just scan people looking at everything, because there's too much. Right? You know, can't keep people in the scanner for 20 hours. And so, you have a problem of how to choose it. And as your selection of stimuli shape what you find, and it's kind of a mess. With audition, it turns out that there's a relatively small

number of basic level sounds that people can recognize. So if you guys all write down-- you don't have to do this, but just to illustrate. If you all write down three frequently heard sounds that are easily recognizable, that you encounter regularly in your life. The three sounds that you wrote down are on our list of 165. Because, in fact, there are not that many different sounds.

And so we did this on the web. We played people sounds. And obviously depends on the grain, right. If it's this person's voice, versus that person's voice, there are hundreds of thousands. Right. But at the grain of person speaking, dog barking, toilet flushing, ambulance siren. At that grain, there's only a couple hundred sounds that everyone can pretty much recognize in a two second clip, and that they hear frequently. And that's really lovely, because it means we can scan subjects listening to all of them. And we don't have this selection bias.

So we basically tile the space of recognizable, frequently-heard, natural sounds. And we scan subjects listening to all of it. OK so here are some of our sounds.

[VIDEO PLAYBACK]

It's supposed to either rain or snow.

[END PLAYBACK]

This is our list of frequency. Most common, man speaking. Second most common, toilet flushing. And so forth.

[VIDEO PLAYBACK]

Hannah is good at compromising.

[VAROOM]

[END PLAYBACK]

So we pop subjects in the scanner, and we scan them while they listen to these sounds.

[VIDEO PLAYBACK]

[CLACK, CLACK, CLACK]

[VAROOM]

Anyway, you get the idea.

OK. So we scan them while they listen to these sounds. And then what we get is a 165 dimensional vector describing the response profile for each voxel. OK. So each voxel in the brain, we say how strong was the response to each of those sounds? And we get something like this. Everybody with me? Sort of? OK.

So now what we do is, we take all of those voxels that are in greater suburban auditory cortex, which is just like a whole big region around, including but far beyond primary auditory cortex. Anything in that zone that responds to any of these sounds, is in the net. And we take all of those, and we put them into a huge matrix. OK. So this is now all of the voxels from auditory cortex, in 10 different subjects. OK, 11,000 voxels.

And so we've got 11,000 voxels by 165 sounds. OK. So now the cool thing is, what we do is, we throw away the labels on the matrix. And just apply math. And say, what is the dominant structure in here? OK. And what I love about that is, this is a way to say, in a very, theory-neutral way, what are the basic dimensions of representation that we have in auditory cortex? Not, can I find evidence for my hypothesis. But, let's look broadly and let the data tell us what the major structure is in there. OK.

So basically what we do is, factorise this matrix. And probably half of you would understand this better than me. But just to describe it, basically, we do a-- it's not exactly independent component analysis, but it's a version of that. Actually, multiple versions of this that have slightly different constraints. Because of course, there are many ways to factorise this matrix. It's an unconstrained problem, so you need to bring some constraints. We'd try to bring in minimalist ones in several different ways. It turns out, the results really don't depend strongly on this. And so, the cool thing is that the structure that emerges is not based on any hypothesis about functional profiles. Because the labels are not even used in the analysis. And

it's not based on any assumption about the anatomy of auditory cortex. Because the locations of these voxels are not known by the analysis. OK.

OK so basically, the assumption of this analysis goes like this. Each voxel, as I lamented earlier, is hundreds of thousands of neurons. So the hope here is that there's a relatively small number of kinds of neural populations. And that, each one has a distinctive response profile over those 165 sounds. And that voxels have different ratios of the different neural population types. OK.

And so, further, we assume that there's this smallish number of sort of canonical response profiles. Such that we can model the response of each voxel as a linear weighted sum of some small number of components. OK. And so, the goal then is to discover what those components are. And the idea is that each component is basically the response profile and the anatomical distribution of some neural population. OK. So let me just do that one other way here. So we're going to take this matrix, and we're going to factorise it into some set of end components. And each of those components is going to have 165 dimensional vector of its response profile. OK. Each component will also have a weight matrix across the relevant voxels there. OK. Telling us how much that component contributes to each voxel. OK.

And then we use, sort of, ICA to find these components. OK. The first thing you do, of course, in any of these problems is, OK, how many? And so, actually Sam did a beautiful analysis, the details of which I'll skip. Because they're complicated, because I actually don't remember all of them. But essentially, you can split the data in half. Model one whole half, and measure how much variance is accounted for in left-out data. And what you find is that, variance accounted for goes up 'til six components. And then goes down, because you start over fitting, right.

So we know that there are six components in there. Now, that doesn't mean there are only six kinds of neural populations. That's, in part, a statement about what we can resolve with functional MRI. But we know that with this method, we're looking for six components. That's what it finds. And so, to remind you, the cool thing about the components that we're going to get out, which I'll tell you about in a second, is that nothing about this analysis constrain those components. There are no assumptions that went in there, right.

So if you think about it, if all we can resolve for the response of each voxel to each sound is, say, high versus low. That's conservative. I think we can resolve, you know, a finer grain of

magnitude of response. But even if it's just high or low, there are 2 to the 165 possible response profiles in here. Right. We're searching a massive space. Anything is possible. Right. And similarly, the anatomical weight distributions are completely unconstrained, with respect to whether they're clustered, overlapping, a speckly mess, any of those things. OK.

So what did we find? I just said all this. OK, so we're looking for the response profiles and their distribution. OK speckly mess, right. Just said that. OK. So what we get with the response profiles is, four of them are things we already knew about auditory cortex. One is high frequency selectivity, and one is low frequency selectivity. That's tonotopic cortex. That's the one thing we knew really solidly.

A third thing we find is a response to pitch, which is different than frequency. I'll skip the details. But we'd actually published a paper the year before, showing that there's a patch of cortex that likes pitch in particular. And that, that's not the same as frequency. And it popped out as one of the components. A fourth one is a somewhat controversial claim about spectral temporal modulation, which many people have written about. The idea that this is somehow a useful basis set in auditory representations.

And we found what seems to be a response that fits that. OK. So all of those are either totally expected, or kind of in line with a number of prior papers. It's the last two that are the cool ones. OK. And the numbers-- actually, the numbers refer to-- never mind. The numbers are largely arbitrary. The numbers are for dramatic effect, really. Component four.

OK. So here's what one of these last two components is. So now what we have is, this is a magnitude of response of that component. Remember, component is two things. It's got this profile here, and it's got the distribution over the cortex. So this is the profile to the 165 sounds. The colors refer to different categories of sound. We put them on Mechanical Turk and had people stick 1 of 10 different familiar labels. And so dark green is English speech and light green is foreign speech, not understood to the subjects. This just pops right out. We didn't twiddle. We didn't fuss. We didn't look for this, it just popped out. And light blue is singing.

So this is a response to speech, a really selective response to speech. It's not language, because it doesn't care if it's English or foreign. So this is not something about representing language meaning, it's about representing the sounds of speech that are present here, and to some extent in vocal music. Pretty amazing. Now there have been a number of reports from functional MRI and from intracranial recordings, suggesting cortical regions selected for

speech. This wasn't completely unprecedented, although it's certainly the strongest evidence for specificity. You can see that in this profile here. Right.

Dark purple is the next thing you get to after the language and the singing. So you get all the way down before you're at dark purple. And dark purple is non-speech human vocalizations, stuff like laughing, and crying, and singing. Right. Which are similar in some ways. Not exactly speech, but it's similar. So that's damn selective. Pretty cool. Yeah? OK. I just said all that.

OK the other component is even cooler, and here it is. OK. Here's the code. Non-vocal music and vocal music, or singing. This is a music selective response. This has never been reported before. Many people have looked. We have looked, it hasn't been found. We think that we were able to find this music selective response. In fact, we have evidence that we were able to find this music selective response. In large part, because of the use of this linear weighting model.

If you then-- I got to show you where these things are in the brain. OK. Running out of time, so I'm accelerating here. We did a bunch of low level acoustic controls to show that these things really are selective. You can't account for them. They don't get the same response if you scramble them. They really have to do with the structure of speech and music. I'll skip all that. Right.

So now we can take those things, those components, and project them back of the brain. And say, where are they? OK. So first, let's do the reality check. Here's tonotopic cortex mapped in the usual hypothesis-driven way. And now, we're going to put outlines, just as landmarks, on the high and low frequency parts of tonotopic cortex. And so, I mentioned before that, one of the components was low frequencies. Here it is. Perfectly aligning with frequency mapping, that this one pops out of the natural sound experiment, the ICA on the natural sounds. And this one is based on hypothesis-driven mapping. So that's a nice reality check.

But what about speech cortex? Well, here it is. OK. So the white and black outlines are primary auditory cortex. And you see this band of speech selectivity right below it. Situated strategically between auditory cortex and language cortex, which is right below it, actually. Not shown here, but we know from other studies. So that's pretty cool. Here's where the music stuff is. It's anterior of primary auditory cortex. And there's a little bit behind it as well. OK.

So we think we were able to find the music selectivity, when it wasn't found before with functional MRI. Because this method enables us to discover selective components, even if

they overlap within voxels with other components. Because our linear weighting model takes it apart and discovers the underlying latent component, which may be very selective. Even if, in all of the voxels, it's mixed in with something else.

So actually, if you go in and you look at the same data. And you say, let's look for voxels that are individually very music selective, you can't really find them. Because they overlap a little bit with the pitch response and with some of the other stuff. So the standard methods can't find the selectivity in the way that we can with this kind of mathematical decomposition, which is really thrilling. I can say one more thing, and I'll take a question.

And the final thing is, we have recently had the opportunity to reality check this stuff, by using intracranial recording from patients who have electrodes right on the surface of their brain. And we've done this in three subjects now. And in each subject, we see-- sorry, this is hard to see. These are responses of two different electrodes over time. So the stimulus last two seconds. So that's 0 to 2 seconds. This is time. And this is a speech selective electrode responding to native speech, foreign speech, and singing.

And here is another electrode that responds to instrumental music in purple, and singing in blue. And so what this shows is, we can validate the selectivity. With intracranial recording, we can see that selectivity in individual electrodes, that you can't see in individual voxels. So that sort of validates having to go through the tunnel of math to infer the latent selective components underneath. Because we can see them in the raw data in the intracranial recording. So this is cool, because nobody even knows why people have music in the first place. And so the very idea that there are, apparently, bits of brain that are selectively engaged in processing music, is radical, and fascinating, and deeply puzzling.

So, you know, one of the speculations about why we have music-- Steve Pinker famously wrote in one of his books that music is auditory cheesecake. By which he meant that, music is not like some special purpose thing, it just pings a bunch of preexisting mechanisms. Like, you know, fat, and sweet, and all that stuff. Right. And so that idea is that music kind of makes use of mechanisms that exist for other reasons. And I think this argues otherwise. If you have selective brain regions, we don't know that they're innate. They're quite possibly learned. But they sure aren't piggybacking on other mechanisms. Those regions are pretty selective for music, as far as we can tell.