Auditory Cortical Processing, presented in partnership with the American Auditory Society

Recorded Oct 2, 2019

Presenter: Sophie K. Scott, PhD

AudiologyOnline.com Course #33919

- [Announcer] At this time it is my pleasure to introduce Dr. Sophie Scott, who is the Director of the Institute of Cognitive Neuroscience at UCL, where she is also Group Leader for the Speech Communication Lab. She is internationally recognized researcher into the neurobiology of human vocal perception and production. Her work addresses both verbal and nonverbal aspects of vocal communication from sound to social meaning. She has published both empirical and theoretical papers on these topics. Thank you for joining us today Dr. Scott and at this time, I'll hand the mic over to you.

- [Dr. Scott] Hello, very lovely to meet you all virtually. So, I'm gonna just start with a very general introduction. So we're going to be looking at some aspects of, initially, quite abstract aspects of how auditory cortex is organized in the primate brain. And then I'm going to be taking that forward into thinking about how this approach helps us understand the kind of picture of processing of sound and of speech that we see in human brains. And how we might be able to start to fit this into a kind of very general sort of approach that might sort of help us think more generally about how the human brain processes sound and how that can help us understand what happens when we're perceiving sounds, what happens when we're producing sounds. So this is just an example of a human vocal tract. And I'm about to show you this vocal tract in action. So if you could play the video for me please. Okay, this is a video that should be playing. I'm just gonna type that into the little box here. Oh no, we might be about to start. Here we go, brilliant. So this is running a structural MRI really, really quickly to give us images of the human vocal tract in action while someone's speaking. So these are just very, very quickly acquired structural MRI images which were taken while someone is speaking. That's the sound that you can hear. So the noise in the background is the sound of the scanner. And what this is just giving you is a glimpse of the kind of complexity of movement and articulation that underlies the speech signal. And when we commonly talk about speech as being a really, really complex sound, it's not unusual to hear it described as the most complex sound in nature. And this is just giving you an idea of where that complexity is coming from. It's coming from an incredibly complex and very specifically evolved vocal tract. And what that's doing is it's taking the sounds made at the larynx and it's actually shaping those with these continual movement of the tongue, the lips, the soft palette, the jaw. And in fact the tongue particularly, if you look at how very mobile, how very greatly that the tongue is distorting. That's really a massive contribution to this sort of very, very finely tuned change to the spectral temporal profile of the sounds that we're making. This is absolutely characteristic of human speech. So we're dealing with this very, very complex sound and what we normally do, historically certainly in my field, is look at that sound and interpret that as being an auditory form of spoken language. And a lot of the, oh sorry. Can I go back to my slides please? Thank you.

Getting trigger happy with the button there, sorry. Brilliant, thank you. So how is that complexity processed in the brain? Well I'm gonna start with an abstract view of a primate brain processing sound. And that's for two reasons. So first of all, although we don't have, um, we don't have clear analogs for human language in other non-human primates, it is the case that other primates do use sound in quite a complex way. It may not be as complex as human speech, but you certainly see elements of sound being used in a highly complex and occasionally with even elements of sort of referential meaning in there, in non human primates. So maybe the picture of the humans being completely different from other primates in this respect may not be totally, totally the case. And it's also the case that over the last 20 years or so, there have been really big developments in our understanding of how sensory auditory cortex is organized in the primate brain. And for context here, we've had a pretty good idea for about 40 years really how the visual system is organized in the primate brain and one of the things that's arisen from this is a suggestion, if not a kind of concrete hypothesis, that what you might see in sensory processing for visual processing in the primary brain can be distinctly different from some of the patterns of processing that we see in other mammal brains. So for example, visual processing in the primate brain is associated with different perceptual streams, at least three different streams of processing. And also, elements of a hierarchy of processing and this is based in it's anatomical framework in an organization associated with the primate brain of sensory cortex into hierarchically organized system. And for a long time, pretty much this was largely studied in the visual system and then developments researchers like Josef Rauschecker, Jon Kaas, Troy Hackett in the 1990's and throughout the 90's started to really explore this in the auditory system. The techniques improved. Josef Rauschecker had the insight that we needed to drive auditory systems in different ways from the visual system. You need more complex sounds, for example, to drive the auditory system in the primate brain. And that started to show us a more general pattern. So I'm just gonna go through this very schematically. And if we imagine that we're looking at the side here of a macaque brain, we're looking at the left side of the brain.

So to the left side of the picture there's the front of the brain. To the right side of the picture is the back of the brain. And you can see these three areas marked out in gray. The darkest gray is corresponding to primary auditory cortex, core auditory cortex that's called, and that's defined frequently as being the cortical fields that are receiving their input from the ascending auditory thalamus and I'm gonna come back to that because it's a little more complex than that. But what we see in terms of this kind of hierarchical organization of auditory, as in the sensory cortex in the primate brain, is we have this core region and then that's surrounded by this lighter band called belt and then another band around that called parabelt. So if we just click through to see that. So we have the core, surrounded by the belt, surrounded by the parabelt. Now, that's just a very general terminology. You could use the same terminology to describe the organization of visual cortex at the back of the brain and elements of somatosensory cortex sitting above these auditory areas. So that's, like, just a general terminology for describing different aspects of the organization of how sensory cortex is organized in primate brains. Now, we zoom in and look in a little more detail within this. What we find through these anatomical studies, these physiological studies from the 1990's, is that there's actually quite a lot of quite specific detail within this, which starts to tell us a little bit about, actually, how these different brain areas in auditory cortex are talking to each other. So what we find is that core auditory cortex can be roughly split into three core fields and I'm gonna come back to this, but for now just think about these like a node in a processing network. And these core fields project out. If we think about it, and I'm only really gonna be talking about projecting out in this lateral direction, so moving away from the middle of the brain. They project out laterally to belt fields and then these project out again to parabelt fields. So that's an element of a hierarchy there. We can also see that the organization of these core, belt, and parabelt fields is built around, sort of, quite a long thin structure. So that core field itself is long and it is thin and it's containing these three core areas within that and they're organized front to back. And the terminology people use for this is rostral to caudal. So rostral, coming from the same sense of the word, like the word rostrum means front, and caudal means tail, that's at the back end. So these three fields are organized this rostral/caudal organization and that's maintained as we move away from the core fields. If we look at the connectivity, what we find is that not everything's connecting to everything.

So very crudely, the core areas at the front end, those rostral areas at the front end of the core shown here in red at the front, are projecting out laterally to similar areas in the parabelt and those are projecting out to similar areas in, sorry in the belt and then the parabelt. So what you're seeing is we have this rostal-caudal organization and to a degree, that's actually maintained in how the connections are coming out from the core and then out from the belt to the parabelt. So you're maintaining elements of this front to back organization. And one of the main things, really, that I'm going to be stressing in this talk is that there's something meaningful about this organization, that it's not just well the structure's long and thin so that's the way that it goes. We then see that from front to, caudal sorry, rostral areas of these auditory fields in the belt and the parabelt, they project, this is from anatomical studies by Liz Romanski, up into the inferior frontal fields. Whereas caudal fields project also to frontal brain areas, but via a more posterior, caudally going directed network and that's projecting forward to adjacent the non overlapping frontal fields. So what we're seeing here, as much as you can see in the visual system, there is very crudely a distinction between information or anatomical pathways that are flowing forward rostrally down the temporal lobe, from these rostral auditory areas into inferior frontal cortex and that's distinctly different from these caudal fields, which projecting via a caudal network, a caudal pathway, into frontal brain areas which are not overlapping with that other pathway. So we've got two roots into frontal cortex from auditory cortex using this very, kind of, general schematic map. We also know from the work with non human primates that there are functional roles associated with this organization. Now if we look within core areas, what you find are tonotopic responses. So tonotopy is, of course, established at the cochlear. It's then maintained in the ascending auditory pathway and all the complexities of that, which I will come back to. And then it's also maintained in the initial representation of sound at the core auditory fields in primary auditory cortex. And in fact, the way that you can define those three core fields within core auditory cort-- Sorry, too many cores here, those three distinctly different areas within the core auditory cortex, it's actually defined by the direction in which the tonotopy goes.

So you see the pattern of low to high, high to low, low to high frequencies flipping as you move across these different areas. So that can be very crudely determined as a sort of, there's a tonotopy here. Beyond that, you don't see much selectivity response in core auditory areas to sounds. It's interested in aspects, that it's responding to aspects of the frequencies of the sounds, but and to other, kind of, contextual aspects of the sounds, but you're not seeing any further selectivity there. As you move laterally, away from these core fields, you start to see more selective responses and this is work by Josef Rauschecker. So as you move away from these core fields, you maintain aspects of the tonotopy, but you also start to see other kinds of selectivity. So you start to see a sensitivity, for example, to the bandwidth of the sound which can be associated with, very crudely, with aspects of the complexity of that sound. And then as you start to move in these rostral directions, what you see are, for example responses in the non human primate brain, they show selectivity to the kind of conspecific vocalizations that you're listening to. So again, this is not as complex as speech, but different non human primates do make different calls in different situations that have a different meaning for their conspecifics and you see selectivity in the responses that's sensitive to these different call types as you move in this rostral direction. Whereas the caudal direction that we mentioned before, that's associated with a couple of different properties of sound. One is where are the sounds coming from? So back here, you'll see cells that are interested less in what the calls are and more in where they are coming from in space. There's also been some suggestion of some quite interesting data showing that in these core auditory areas you also see a response to tactile sensation around the face and around the mouth. So maybe associated with the sort of sensory information that you might use to guide the production of a sound yourself and I will come back to this.

So we don't have to have an answer for exactly what's going on here, but there certainly does seem to be a roadmap. We've got this network which is showing the rostral/caudal organization and then with different kinds of projections to frontal brain areas and if we look at different sorts of sounds, those do seem to drive or be associated with different patterns of processing in these non human primate studies, such that we're seeing selectivity for sounds more generally as you move forward in this rostral direction, whereas the senstivity to where the sounds are coming from and potentially to motor control aspects of producing the sounds yourself in these caudal fields. So this has been quite an interesting framework to take into looking at the function limiting studies of intact human brains and that's been very useful to us. So historically, this is a subject that we have studied by working with patients who have had some kind of brain damage and we have learned a great deal from working with patients who have, for example, sensory aphasia or production aphasia, by being able to interrogate the relationship, for example, between patterns of brain damage and the sorts of problems people have with understanding speech or with producing speech. And of course the classic work of Broca and Wernicke are in this area are very well established. However, we develop these new techniques in the 1980's and 1990's that let us start to take photographs of intact brains in action, as it were, using things, initially like positron emission tomogrophy and then more recently functional magnetic resonance imaging and of course EEG, ECoG, and MEG studies. These have started to let us ask questions about what's going on in a brain where nothing has gone wrong and we're trying to, sort of, capture aspects of the quotes normal function. And it's been actually quite useful to have this framework from the non human primate. So very, very generally, the way that we started to think how we could characterize these different pathways that we're seeing and how that we could take it into the human brain. These rostrally directed pathways, showed here in pink, those seem to be associated in certainly the non human primate brains with what are those sounds. So that might correspond to understanding speech in a human brain, whereas these caudal pathways they seem to be involved in both the spatial location of the sounds and also in, sort of, sensory motor guidance perhaps of producing that sound. That seems to be associated with some qualitatively different way. So perhaps, where the sounds are, how the sounds could be made and that's very, very generally is wrapping on to quite similar aspects of the streams of processing that have been described in the visual system.

So it does seem to be that there's some general coherence there across the senses, as well as potentially across the different groups here in non human primates and humans. But, we didn't have a very good understanding of what could actually be driving this, how would the brain be dealing with this information. And if it is dealing with it in an anatomically distinct areas as it suggests that there's some sort of computational difference there, but what might that be? Well, that has become a little bit clearer over the past few years with developing studies. So what we wanted to know was, okay we can see these streams of processing, but are there any salient organizational principles that underlie these streams? Are they doing anything qualitatively different when they process this information about the sounds that we're hearing? Can we characterize computational differences that we could actually use to describe and delineate differences between these pathways. And of course, can we ever come up with something that's a bit less tied to one particular kind of group of sounds? So I tend to think of myself as being somebody who works in speech, but of course I'm always having to tackle the fact that this is a sound and it makes no sense to work in this area or not think about the acoustic properties of the signal. But even then, I tend to, you know, you could have a whole career studying speech and never, for example, think what would happen in the brain, these same areas, if the brain wasn't processing speech, but was processing music or some other kind of class of environmental sounds, some other sort of sound.

And I think it is quite interesting that there are actually lots of different theories out there. People have written about processing emotional sounds within these networks, processing of speech as I had written about, processing of language, processing of music. It's always very domain specific. People have a set of data about music and they look to see how we can, sort of, map this onto the auditory brain. And something that I and my colleagues, Cesar Lima and Kyle Jasmin, were interested in doing was asking the question, "Can we take a step back and actually come up with, can we just determine any of these computational differences that might underlie these apparently different perceptual processing scenes?" And actually, "Would that help us start to get a grip on what a more domain general theory of auditory processing might look like?" So that's really what I'm going to try and go into now. And to give you the answer at the top, there does seem to be at least one way that we can delineate a difference between these rostral and these caudal auditory processing streams, does seem to be in terms of how these streams, how cells, brain areas within these streams, respond to sound in terms of the temporal response characteristics of those brain areas. Very, very crudely, as we move in a rostral direction as we move forward down the temporal lobe, what you find are cortical responses that respond to sounds, potentially any sounds, in a way that is slow and sustained so that what pathway seems to be associated cell responses which responds slowly to sounds and then stay activated. Whereas, these caudal regions, this how or this where pathway, that's much more associated with cells that respond quickly to sounds and in a very transient way. They die away almost immediately. So what could this mean and where would we have got this data from? Well, there's been a few interesting papers that have fed into this. So first of all, something that it is very important to bear in mind and something that if you work in my area, where we're looking mostly at cortical activation it's easy to forget, but of course there is an enormous amount of pre-processing of the sound in the ascending auditory pathway before we've got to the signal that we're seeing represented in primary auditory cortex and beyond. So I'm not going to go into this in great detail, but just to bear in mind that the sort of signal we are already dealing with, by the time it gets to the cortex, has been highly processed. So even within one synaptic connection from the cochlear, by the time we get to the cochlear nucleus, we are seeing a large number of parallel acoustic features being detected. There's already been a lot of some spectral detail, there's some kind of compression of the signal and we're seeing this tonotopy represented across all of these parallel acoustic features that are being detected.

So we've already seen the signal being decomposed in ways that seems to map in, for example, things that are gonna be useful for knowing that the spatial location of the sound, but also aspects of the pitch of the sound. And then as we move up the ascending auditory pathway, a lot of this information, these parallel streams are maintained. We see information being decoded about the spatial location of the sound. We see the signal being cleaned up, reverberation is being sorted out at, at least the cortical levels. And by the time we get up to the thalamus, so where we've got the medial geniculate body, but also other thalamic nuclei, you've also got the possibility the information being integrated with other sorts of sensory information We've also got a few other characteristics as we move in this direction. So as we move away from the cochlear, we get this decrease in temporal resolution. We have this reduction in the spectral resolution. So you're seeing the information processed to a great degree and also, um, there's a distinct possibility that the information that you might be representing cortically could already be information that's been, kind of, computed in the ascending auditory pathway and now is being represented cortically into a plastic, more flexible way. So I'm not going to go into this in any more detail, but it's just worth bearing in mind the kind of information that is in there. And then it also turns out that there is a rostral/caudal aspect to the connectivity of the ascending auditory pathway and where it's speaking to, where it's connecting to these core fields. So, just to forgive the slightly crazy wiring diagram here, this is redrawing data collected by Brian Scott and published in 2017. And what you have to imagine here is we're looking at the core auditory cortex fields moving forward down the temporal lobe. So A1 is right at the back end of core auditory cortex, R is in the middle, RT is at the front, RTp is a beltfield just in front of the core, and STGr is a parabelt field.

So we can imagine a line moving forward from back to front in this caudal going rostral direction. What you see is a big difference in how the thalamic nuclei are projecting to these auditory areas. So the medial geniculate body, which is made up of several sub components, we're seeing the MGV, the ventral medial geniculate body is projecting primarily to A1 and to R. And then as you move forward from RT, you start to see that the medial geniculate posterior dorsal nucleus starts to contribute more to the input. RTp is getting more of its input from other thalamic nuclei, such as the, I'm furiously looking at my list at the medial pulvinar and the SG Lim body. So by the time you're at the front end of these auditory fields, STGr, you're actually getting almost none of the input from the quotes auditory thalamus and more input from these other thalamic nuclei associated, for example, with more visual processing. So there does seem to be, not only great complexity of course in the ascending auditory pathway, but there does seem to be when we look at the thalamus, some rostral/caudal organization in what information is actually getting into the auditory core fields, which of course again is represented in this rostral/caudal distinction. And then if we look within the core fields, and this is another paper by Brian Scott, so this is again just a snapshot of some of this connectivity. Excuse me coughing at you. What you can see here again, you can see A1, R, and RT so those three core fields delineated there. A1 is the caudal field, RT is the most rostral core field. If you look at the connectivity within these core fields and where they're talking to, you see that robust rostral/caudal distinction is maintained. And what that means is that A1 is projecting backwards and forwards to R and R is projecting forwards to RT. RT is then projecting forward to the RTP, that parabelt field and then that's projecting forward down the temporal lobe. And very, very crudely, you are seeing the same pattern of rostral/caudal distinction being maintained in much, I showed you a rough cartoon of this earlier, but this is showing you a much more detailed image of how that information is being represented. And you could imagine an organization of auditory cortex where there was much more interconnectivity. A1 could be talking to RT, RT could be talking to CL. There could be much more interconnectivity and this is not coming through. It does look like when we start to get into some more and more detail, the rostral/caudal representation which is being informed by the ascending auditory pathway is also being maintained within the core fields. Now a persistent critique of this approach, and it's not an inaccurate criticism at all, is to say, "Well, what you have here is a long thin structure." "It's long and thin, you couldn't have it organized in any other way because actually it's sitting on top of the transverse temporal gyrus, it's sitting at the top of the temporal lobe tucked away inside the Sylvian fissure and it has to be long and thin, it can't be any other shape" "Visual cortex can be a different shape because it's sitting in literally a different brain area."

So this rostral/caudal organization has always had the suggestion that it could just be, well that's the shape that it is for some other anatomical reason, there's no necessary properties that arise from this. And I think when you look at this kind of pattern of the ascending auditory pathway and other thalamic nuclei and how they project to these rostral/caudal fields differently, I think that does suggest there might be functional differences. And when you look at this pattern of connectivity, again that does suggest there could be functional differences, but if we look in detail can we actually see those within here? Well, this is taking some work again from Brian Scott in 2011, again suggesting that there could be. So as we move forward, again this is just showing you that caudal to rostral direction, looking at different auditory fields. Brian Scott has shown you pick up. Now, instead of saying how these brain areas connected to each other, how these cells connected to each other, he's saying, "How do they respond to sound?" What you find are differences. So the two graphs here, shown in red, are showing us how, sorry I've got this on the wrong side. So if we look at the two panels on the left, we're seeing how quickly these cells in these fields respond to sound and what you find is that in caudal auditory areas, so that's A1 shown in yellow, what you get is a faster response to sounds than you do in the rostral field laying just in front of A1. So just to go back, we're comparing these two fields here, A1 and R, so the one's right at the back of these core fields. Now these differences are not big, but they are robust in their significance. And what this is showing you is that when a sound happens, there is a faster response to that sound in A1, that most caudal field, than there is in the immediately adjacent core area, R.

Similarly, if we ask a different question about temporal response properties instead of saying, "How quickly does this cell respond to sound?" We say, "How quickly do these cells or how well do these cells entrain to amplitude modulation?" What we find is that in the caudal area, A1, what you find is a more accurate tracking of amplitude modulation, going from slow through to fast, than you find in the immediately adjacent R field, the rostral field, the next core field along, which actually starts to maximize its response at much slower amplitude modulations. It cannot track the faster amplitude modulations. So just on two temporal measures here, we are seeing differences within these core areas just in terms of how quickly those respond to sound. And to take completely different snapshot on the same question of temporal responses, this is now looking at data from humans and this is a study by Liberty Hamilton that came out last year and she's been working with what's called ECoG data. This is data collected, it's electrical data that's measured from the surface of the brain when people are having preoperative investigations for surgical mapping prior to having epilepsy surgery. So this is measurements, it's not quite as fine as inserting an electrode into the brain, but you've got the electrode on the surface of the brain so you've got very detailed and very, very fast, temporally precise measurements of how the brain area, the cells directly underneath that electrode, are responding to, in this case, sounds. And Liberty was working with Eddy Chang and Eddy Chang has published a large number of papers over the past few years asking some really elegant questions about the processing of speech and sound in auditory cortex using these ECoG measures. What they did in this study, was simply take all of the ECoG measures they've ever collected and used machine learning just to classify not what those cells responding to, those electrodes responding to, instead they said what other response characteristics measured at these electrodes? Is there anything systematically different in terms of how those electrodes respond to sound? Well the brain areas beneath those electrodes. And this is what they found, what they find is that in rostral areas, like the two panels shown at the top in pink, those show fast, transient responses to sounds and in fact it could be any sound. Most of what they worked with is speech, but the speech wasn't all intelligible. And whatever the speeches or the sound people are listening to, what they see is these very fast responses that decay very quickly. In contrast, in rostral brain areas they find, rostral/caudal, sorry rostral auditory areas. What they find are brain responses that are slower to any sound and much more sustained and this does seem to speak to different properties in terms of the, well the computational aspects of what's happening underlying this.

And it's very tempting, although it's very hard to map from single cell recordings in non human primates to these electrode recordings in awake humans, it is interesting that we're seeing the same rostral/caudal distinction. The caudal areas are responding quickly, the rostral areas are responding more slowly. The caudal areas are dying away fast, whereas the rostral areas are remaining engaged on a much longer time scale. Now for example, if you were to take a much, much, sort of, messier method, if you like, for looking at brain's electrical activity, so the world of looking at EEG and ERP studies, what you find if you use, ERP, so measuring the brain's electrical field generated by brain activity by putting electrodes on the surface of the scalp. People have done lots of studies looking at the time scale of semantic processing or syntactic processing using this, sort of, classically. You have people listening to sentences or reading sentences where there is some kind of violation, for example, I have a semantic violation or syntactic violation. Now, the time scale for the brain responses to those violations is slow. It's on the order of hundreds of milliseconds. It's much more consistent with the slow sustained responses that Liberty Hamilton has described in these rostral brain areas. In contrast, if you look at studies where people are producing speech or they're producing sound, that does seem to be something that has a much finer temporal resolution and can enable much finer temporal processing than necessarily recognition networks. So here's just an example of this. There was a study by Bruno Repp from a couple of years ago when he measured people's ability to be able to track interval differences in sequences of just click sounds. So people would here a and they're having to listen for a difference, a change in that interval, and then when they notice that change, you establish how big a change it has to be before people can report it. And that gives you a, you know, measure of their temporal sensitivity. If instead of having people listen to the clicking sequence and let them tap along to it, what you find is that their sensitivity to changes in the intervals is much better and also they will frequently adapt to an interval change that is smaller than one they could actually consciously report on.

So it certainly suggests that if you let people engage motor responses, this how network when people are responding to sounds, they are certainly showing better temporal resolution of those sounds. The motor system maybe just has better access to the temporal resolution and maybe because it is depending more on these caudal pathways that have more accurate temporal responses. So we've got this very crude distinction. I don't want to imply at any point that this is the only way in which I think we're going to be able to find differences between rostral and caudal brain areas. But if we then take this as a framework, does this help us understand any of the brain activities that we have been recording using function imaging studies like PET and fMRI over the past few years? Do we seeing anything that starts to map onto this framework and anything that might be associated with, you know, perhaps something more associated with these sustained responses in rostral areas and fast onset responses in caudal areas? So just to summarize, what we're seeing here, very generally, is in caudal auditory fields we see neural responses, so the cell responses in cells, are quick, they have short response latencies. They respond in a transient way. They can track fast and slow amplitude modulations. They receive somatosensory input so it's a good place to have some kind of integration of sensation of sound and also of somatosensation associated with making a sound. We think, and I'll show you some evidence for this, this might be why these brain areas are particularly important for guiding online sound production, so what I'm doing now, speaking. It seems to be particularly useful or important for processing soundless actions and also of course the spatial computations, the spatial information associated with sounds needs to be represented accurately in time as well as in space. Whereas these rostral brain areas, they have a longer response latency. They show sustained responses to sounds. And certainly in terms of computational processes, this would be consistent with a neural response which is responding both to incoming sensory information and also to feedback of higher order representation of information in the system so that you're seeing the potential in this pathway for some sort of processing based not only what's coming in, but also on expectancies, prior knowledge, semantic knowledge for example. It's only really accurate at tracking slow amplitude modulation. It's not accurate at tracking fast amplitude modulations. It receives visual input. It seems to be important in recognition systems and of course we know recognition does not perceive just by using incoming sensory information. You use existing knowledge to help guide this and understand this. It connects the semantic system. It is potentially capable of processing multiple auditory streams and I'll show you examples of that. So now if you move on to the empirical side of the talk, evidence for these where/how pathways and these what pathways. This is a very old study where what we were doing was comparing different kinds of intelligible speech. So here we've got intelligible speech

- [Example audio] They're buying some bread.

- [Dr. Scott] And then we've got that same speech, but now we've got it noise vocoded using-- technique. And what we were looking for, this is a very old study now, but what we we're trying to do was identify brain responses which would be activated by speech you understood regardless of what it sounded like. So we wanted to know about a brain area that cared

- [Example Audio] They're [Dr. Scott] as much about-- buying some bread. As about. And because what you're always looking at with functional imaging studies is something that has a relative pattern of activation we used it to use baseline conditions. So we've got these two spectrally rotated conditions. which are acoustically complex, but which do not, that you cannot understand, that you can't process those as speech. And when we do this, what we find is this area in yellow is showing you a response which is activated for the speech and it's activated for the vocoded speech, but it is not showing a sensitivity to those two baseline conditions. And in fact, as you move away from primary auditory cortex forward down the temporal lobe, you actually see the response in the brain level become more and more selective for the speech that you understand regardless of what it sounds like or it sounds like who's talking. Now, this was really interesting because this was actually the first unambiguous demonstration of lateralized response to speech and the intelligibility of speech, which we knew clinically had to be out there, but it was actually hard to demonstrate. We had to get our baselines right.

Now, this is also very interesting because it was lying in these rostral areas and the dominant view at the time was that there should be something sitting at the back of the temporal lobe and then we saw these monkey studies that were coming out and we realized, well maybe that's what we're seeing. We're seeing this what pathway recruited. Can we see it for any other aspects of speech processing? Well first of all, of course, you need to replicate this and actually this is a study that has been now widely replicated. We also went in and asked the question, effectively the same question, but now instead of using connected speech we used single phonemes like And when you do this, which I can honestly say without a shadow of a doubt was the most boring experiment I've ever been a participant in, you still see selectivity running within these anterior pathways. If you look a the plus at the top here, that's showing you activity of speech over two different baselines, click sounds and signal correlated noise. And even though listening to is not speech in any meaningful sense, you've just got these isolated phonemes, you are still driving that anteriorly directed pathway so it's responding to speech and it's showing a left dominance. Now, I think it's probable fair to say that if you listen to you might hear cat or cot or cut and that's because, of course, the recognition isn't only proceeding in this ascending way. Your brains going to be trying to hear speech there. It's picking up a suggestion that there might be intelligible speech there and I'm certain that's contributing to this activation. But in fact, I think that would actually still be consistent with these recognition pathways involving both input and existing knowledge about the world to help you decode what's going on. We see the same brain areas recruited if we look at syntactic structure and several other labs have shown this.

So whatever is happening within this pathway is not only processing semantics, it's processing syntactic information. And we're also seeing sensitivity to non speech sounds. So this is a study where instead of using intelligible speech we used intelligible non verbal vocalizations, so things like screams and laughs. And what you're seeing is a very, very similar pattern with the dominant response at the front end of the temporal lobe and that gets, sort of, more selective as you move away from primary auditory cortex. The only difference here is that the peak is on the right, not on the left. So you're seeing an element of hemispheric asymmetry here and, really, we do see this quite markedly throughout these studies, particularly if you work with speech that's intelligible and other aspects of the voice. And the way that this seems to boil down is there is a selectivity on these rostral areas on the left for linguistically relevant information, phonemes, semantics, syntax, whereas other vocal characteristics, emotion, speaker identity, intonation, that seems to be strongly right lateralized within these rostral areas. And in fact, this is another study where we expressly looked at this. So this was a study where we varied speech intelligibility with vocoding and we varied, along with my colleague Stuart Rosen, whether or not the speech had a naturalistic intonation profile. I'm afraid I don't have an example of this so you'll have to tolerate me impersonating it. So people overheard a sentence like, "They're buying some bread." or they heard the same sentence with a flattened pitch profile, like, "They're buying some bread." Now, I was quite worried when we ran this study because I thought we're gonna get loads of activations that really weird pitch condition because they're talking like this is very strange, it's very salient, it really leaps out, but that's not what we found. We found a strong right lateralized response specifically to the natural pitch profiles, relative to the flattened pitch profiles and it was strongly right lateralized. So again, this can't be the end of the story because, of course, intonation and meaning strongly interact when you're actually listening to somebody, but there clearly is some basic preference for the right temporal lobe for processing that speech rather than intonation profiles.

We also looked at some of the ways that facial information can contribute to speech processing. So in this study, we had people listening to noise vocoded speech and we deliberately used low levels of noise vocoding so you had to use the speaker to help you understand it and the speaker was always first It was always a talking face along with the noise vocoded speech and there were four different talker conditions. So she's either looking straight at you or she's looking down or she's looking away or her eyes' blanked out. And if you look specifically at the condition where she's looking directly at you, what you find is that drives greater activation in left rostral areas. Now, this is really interesting because the eye gaze did not effect how intelligible the speech was. Those intelligibility in all these conditions was the same because, primarily it's the mouth moving that's helping you do that. However, when the talker is looking at you, that suggests that there is some other meaning to this. This speech is more salient because it's directed at you and that seems to enhance the signal in these left rostral auditory areas. So you can drive this visually, not just acoustically. And I made reference before to the fact that there's possible the case that these rostral fields process multiple streams of sound and that's because we've done several studies now where you've had people listening to target speech where we play them different sounds which are competing with the target speech, different sorts of masking sounds.

And when you do this, you can find brain areas that show a general response to masking and those are actually shown here in blue. So when you're listening to speech against any other competing sound it's a harder task to do and you see these attention networks shown in blue recruited. As soon as the competing sound is another talker, you also start to get more activation run rostrally down the left and the right temporal lobes. Now, that is almost certainly associated with processing that unattended speech because, of course, we've known for decades that when you're listening to a talker and there is somebody else speaking, there is the possibility of processing information in the unattended speech, particularly if it's semantically relevant. Of course, Douglas Brungart has shown what we call informational masking does seem to strongly point to competing for resources at a central auditory level and that's what we're seeing here. I think we need more studies to get to grips with this, but I think what this suggested that these rostral pathways are also capable of processing multiple streams of auditory information and, of course, selecting between because you can switch your attention from one to the other.

So there's a great deal more to know about this. Interestingly the same is also true when you're speaking. So this is showing you the exact same effect, but now people are speaking rather than listening to an unattended speaker. You get the exact same dominant part, so soon as you are speaking and there is somebody else talking you see all this activation in the temporal lobes associated with processing that unattended talker. So again, the unattended speech is being processed very, very thoroughly in auditory cortex, running sown these rostral pathways. And this doesn't stop there. So, of course, we don't just process speech as a sound. Speech is integrating with the widest semantic network and if you do something like simply add in a larger number of participants, so here we've got people from two different language groups listening to speech against a baseline sound, the English and Mandarin speakers. You're seeing these anterior auditory fields on the left recruited for intelligibility, but you're also seeing rostral temporal pole and brain areas on the bottom of the temporal lobe. Brain areas which are associated very strongly with the semantic processing of sound. So this is a model semantic representation that would be driven as strongly if you were reading that with braille or if you were reading words. So we can see a very strong pattern of these sustained, slow responses in the how/what pathway being very, very strongly linked to aspects of recognition certainly within speech and I'll come back to read some examples of non speech afterwards. Well of course, we don't only listen to speech. We also produce sounds and this is quite an old study now just looking at the brain networks recruited when you are talking in the scanner. Lots of bilateral stuff to do with motor control. We also see three fields which are highlighted here in yellow which are left lateralized and remember speech and language are largely left lateralized in the human brain. And particularly draw you attention to this little blob here because this is very interesting. This is a blob of auditory cortex which is very, very robustly activated when people talk. Of course, when you talk you're hearing your voice. Maybe it's just responding to the sound of your voice, but in fact this is another study where we had people speaking, but also we had people mouthing silently and you still see it activated. So this actually seems to be a part of auditory cortex which is activated when you speak, even if you speak silently. If you're guiding, you're articulated, you're sound producing effectors using some information about sound to do so. What are the implications of this? Well first of all, of course, this is sitting slap bang in those caudal auditory areas so that's very interesting to us. It seems to be potentially part of this how pathway.

And this is another study where we look to this in more detail. And here we've got people speaking, we've got people mouthing silently while they hear someone else speaking, or we've got people just listening to another talker. So we've got three conditions here. The thing I want to draw your attention to is twofold. If you look at the regions in blue, those are activated when you're speaking and when you're mouthing silently. So that's really a tremendous amount of recruitment of posterior auditory fields. When you are using your articulators, whether or not you're making a sound, interestingly rostral auditory areas which you would normally you using to listen to other people which is shown here in yellow are actively suppressed when you are speaking. So when you yourself are talking or mouthing silently, you are actually shutting off those brain areas that would show a response to decoding speech of other people, possibly because this is just the way that the brain deals with self generated sensory information. You see something very similar in the somatosensory cortex if you touch yourself, So you're sort of discounting that signal. So we're seeing a rostral/caudal pattern here in auditory cortex during speech production or silent mouthing which is showing a rostral response which is a suppression during articulation whereas these blue areas, these caudal areas, are strongly recruited. What might be a functional role for this? Well, here we've got-- can I just play this video? This is somebody speaking under delayed auditory feedback and what I want you to listen to is how his voice is being affected by hearing his own voice come back to hi--

- Understanding the constraints on Kennedy, Hanner and Hesper-- berg wanted a commission to exert counterpressure by having special access to the white house through a liaison. Kennedy said that Ha-- Harris Walford who me and made a specials-- a full time specialist assistant on civil rights was already on the job, which was false. But Hapner and Hesburgh responded to the Walford-- It wasn't--

- [Dr. Scott] So I'll have them go back to my slides now. So you could hear that he was getting very stuck on the starts of words and delayed auditory feedback is very interesting because it seems to trip people up a great deal and make people become a lot less fluent. Interestingly, if you have a developmental disfluency, like a stutter or a stammer, it tends to make you more accurate. So what we did is we took people who didn't make mistakes like this and we scanned them speaking under different amounts of auditory delay and what you find is as you vary the amount of the delay, you make it harder and harder to speak, you see brain areas recruited as the task gets more difficult, but all of them, interestingly, are in these caudal auditory fields. So we have a speech production task which we are making difficult progressively and the harder it gets the more you are seeing these caudal auditory areas recruited. So the detection of some mismatch between what you think should happen when you speak and what happens when you speak and the compensation for that really seems to be coordinated in these rostral areas. Sorry, in these caudal areas. And we have a paper under review at the moment where we've done a meta analysis. My former Phd student Sophie Meekings and I, Sophie's done a meta analysis of a wide variety of studies that have modulated some aspects of what your voice sounds like when you're speaking and then has looked to see for the brain areas that are common across all these studies. So delayed auditory feedback, altered pitch feedback, altered spectral feedback and in fact, they're all showing responses sitting at the back end of auditory cortex, these caudal fields. So when you are changing your voice, when you're speaking, you recruit these fields, and when you change your voice in some way to adapt to the situation around you, you are doing the same thing, you are recruiting these caudal fields. And-- I think this might have appeared in the wrong place. Is it possible to go back to the next slide? Thank you. No worries. Okay, so I have not put that in the wrong place.

So I've just come back to this video in a second. So what we can see is that certainly in the case of speech production you're seeing a whole network of motor systems recruited when we speak, but there does seem to be a very important role for these caudal auditory areas in the production of speech and in the changing of the constraints. When you need to change your voice when you're speaking you see these systems recruited differently. And in fact, we've seen this across other situations. So when you are changing your voice to align it with somebody else, you see these caudal regions recruited as well. Now of course, one of the things about the motor control of speech is it has to happen, like any motor act, extremely quickly. You cannot wait to recognize your own speech to work out if you've made an error or not. If your task depends on timing then you need to get that right in the moment and that's probably why, potentially, these caudal areas with much faster auditory responses are particularly important for the coordination of making a sound in the world with your voice. Interestingly, it's almost certainly not limited to speech. This is something that you also see, for example, with people who are playing musical instruments and you introduce a delay. So I just want to finish with another couple of extra points to add on top of this. So we can see there probably is something quite systematic in the rostral/caudal organization of primary auditory cortex and these surrounding fields and these pathways. We can see how recognition processes that run forward down the temporal lobe and these rostral fields, those do seem to be very comprehensibly involved in the recognition of different kinds of information from the speaking voice with the most hemispheric asymmetries within this, but that has a distinctly different profile from these caudal responses which seem to be very, very important in coordinating the sound of your voice and changing the sound of your voice depending on the context that you're in. But of course, we don't only speak and there are other kinds of things going on. So I've got an example here, if I could have this first video again please, thank you. So this is somebody beatboxing in the scanner. Now I generally tell people speech is the most complex sound in nature and then I met beatboxers and I realized that we're doing the bear minimum, really, when we speak to each other. There's a great deal more that we can do. So he's making at least two different sounds most of the time here. Okay, if I could go on to the next slide please, thank you.

And what we did was, we wanted to look at this, kind of, vocal expertise and musical expertise. So we took a group of people who were expert professional beatboxers, like that guy there, and we also took a group of people who are expert professional electric guitar players. And we deliberately chose both these groups because they tend not to have a formal musical training. They're making their music in very different ways and they're both groups of experts, they are professionals so this is an expertise. And we compared them to non musicians listening to novel examples of beatboxing and novel examples of electric guitar playing. And what we find is that as soon as people are listening to the instrument they can play, you see a different network created. So you can see the guitarist in the top panel in the middle listening to guitar music. The beatboxers in the lower panel on the right listening to beatboxing. But interestingly, there's a very strong expertise effect which is almost exactly the same for both groups.

So when you are listening to the thing that you can play, you recruit a very similar network across beatboxing and across electric guitar playing relative to non musicians and the rest of the time you look like the non musicians. And of course, the interesting thing here is twofold. A, this is a largely a left lateralized network, often sitting in these caudal areas and going around, projecting up to motor and pre motor responses, but of course we're all experts in speech production and perception and we don't think of ourselves as expertise because we spent a huge amount of our lives learning to speak and to understand the language that's spoken around us. How much of the stuff that we're seeing in the response to spoken language actually used to do with this expertise? And of course, the other thing is the words are not the only things we say and we do have different sorts of reactions to sounds. So I just want to finish with this example. So this is somebody on the radio in the UK who hears something that starts to make her laugh. Okay, so this is a live news broadcast. The pitch of the voice is shown below in blue.

- [Male radio Broadcaster] popular replacement is now being dismissed with the army's popular chief of staff, Jack Twat, taking over.

- [Woman] Jack Twat.

- [Female Radio Broadcaster] The 40 foot sperm whale, which was stranded in the Firth of Forth for more than four days is now thought to be swimming towards open waters again. It freed itself late last night. Marine experts are hoping to establish this morning whether the whale is finally back at sea.

- [Male Radio Broadcaster] Good luck to the whale. Ten past eight is the time an investigation is underway at the ma-- So of course, words aren't the only things we say. She's in the studio and actually the guy coming down the line has to say a silly name, I don't know how it works in the US but Jack Twat is, of course, a rude name in the UK and he just goes for it, Jack Twat. Someone back in the studio leans in to the woman about to read the news and goes, "Jack Twat" like this at her. And they're doing that specifically to make her laugh and she's okay at first and then she starts to laugh. You can see the effect it has on the pitch of her voice. By the end, she can no longer speak and she's just making, sort of, squeaking noises because now on live radio she is laughing. Now, laughter is a very interesting example of a behavior that we produce a great deal in social interactions and it's frequently interleaved with speech, but it has quite different, sort of, behavioral characteristics associated with it. So for example, laughter is highly behaviorly contagious. Very often a laugh you produce is happening entirely just because somebody else laughed. And so this means that, you know, speech itself is of course a social phenomenon, but laughter, for example and then speech, is highly associated with this pattern of, you know, actually catching a laugh from other people. And we can-- sorry let's skip past that. You can actually see that if you look at the neural responses to laughter.

So this is people listening to emotional vocalizations and you see very strong orofacial mirror response when people listen to emotional vocalizations like screams and laughs. That's shown in the top in green. But if you actually mapped this out by the valence of the sounds, how positive and negative they are and their arousal i.e. how arousing the sounds are, the more arousing they are the more positive they are, the more they drive this orofacial mirror system and that's shown below in pink and blue. Such that when you listen to laughter, you get the biggest orofacial mirror response of all and that's almost certainly associated with this contagious use of laughter. So I think it's also very important when we look at the neural systems that are recruited by vocalizations, not only to think about, you know, are you hearing a communication sound, but actually what the use of that would normally be. And it's not always the case that it's being decoded in the same way frequently for sound that is perceived in this contagious way. There'll be some other aspect of behavior required of you and, of course, it's important to bear in mind we learn to laugh contagiously. Babies don't laugh contagiously, this is all a learned behavior. So to finish with these networks, we've got strong empirical evidence for different recruitment of brain areas when you are listening to speech and you are for understanding recruiting the what pathway more. When you are speaking and when you are changing your voice, adapting to the environment, that is recruiting these how pathways more. I haven't really talked about the where pathway at all, but it's possible, of course computationally, that it's not particularly distinct from how. If you're acting on the world you need to get your actions accurate in time and in space. So speech and voice processing is resting on a rostral/caudal auditory network and it will recruit those differentially depending on what you're doing. And if you're listening to speech you might recruit these contagious responses less than if you are listening to laughter. There is substantial hemispheric asymmetry, particularly for aspects of speech in these rostral areas and that can extend to frontal areas, certainly in speech production.

Again, it's a heavily left lateralized task. And we can also see these networks recruiting, depending on what the sounds are, into mirror systems. So you're seeing aspects of mirror responses when the musicians were listening to an instrument they could play and we see mirror responses when humans listen to sounds they normally produce contagiously. So the behavioral meaning of sounds will affect the networks that you see recruited. And I think we also need to bear that in mind when we're thinking about speech. So I'm gonna stop there. Thank you very much for listening and if we have any time for questions, I'm very happy to take some now. Okay, there's a question in the queue. "During speech production is activity suppressed in the caudal or rostral fields?" During speech production activation is being suppressed in the rostral fields and during speech production activity's increasing in the caudal fields. So you're recruiting the caudal fields to help you speak and you're actually turning off the rostral ones which you would use for understanding someone else's speech. And that's probably because when you actually produce the speech act, you've already planned it and you know what you want to say and by the time you're speaking you're already planning the next thing you're going to say. So it's more important, your attentional resources go there. So very frequently, actually, in speech production people often do not actually correct errors that they make when they're speaking, even if it really matters when they're talking. So that's probably because you aren't decoding your own speech for meaning and certainly you turn those brain areas off. The next questions is, "Which connectivity in the primary auditory field is preserved?" Well, in the primary auditory fields, remember you've got three, the A1, R, and RT. And they are connecting in this rostral-caudal direction. So A1 is primarily connecting within the core to R. R is connecting to A1 and to RT. And RT is projecting mainly to R. So you've got this rostral/caudal within the core, the rostral/caudal dimensionality is being mostly preserved in how they talk to each other as well. And then they project out from that to lateral fields in the belt and that's also mostly maintained in the rostral/caudal organization. Okay, so I'm going to wrap up there. Thank you very much for your time and I's just like to take this opportunity to say thank you very much to all my collaborators and people in my lab for this work, thank you.