OPINION

Understanding rostral–caudal auditory cortex contributions to auditory perception

Kyle Jasmin, César F. Lima and Sophie K. Scott

Abstract | There are functional and anatomical distinctions between the neural systems involved in the recognition of sounds in the environment and those involved in the sensorimotor guidance of sound production and the spatial processing of sound. Evidence for the separation of these processes has historically come from disparate literatures on the perception and production of speech, music and other sounds. More recent evidence indicates that there are computational distinctions between the rostral and caudal primate auditory cortex that may underlie functional differences in auditory processing. These functional differences may originate from differences in the response times and temporal profiles of neurons in the rostral and caudal auditory cortex, suggesting that computational accounts of primate auditory pathways should focus on the implications of these temporal response differences.

The primate visual and auditory perceptual systems are equally complex, but approaches to untangling their complexity have differed. Whereas models of visual processing have often been examined in a domain-general way (by measuring neural responses to basic visual features, for example), models of the auditory system have tended to focus on specific domains of auditory processing, such as the perception of intelligible speech and language[1–3], the perception of linguistic and emotional prosody[4,5] and the perception and production of music[6,7]. Studying these specific domains has proved useful for determining the functional properties of the auditory cortex, and it is arguable that beginning with such approaches was in some ways necessary. For instance, the functional organization of the macaque auditory cortex into a rostral 'recognition' pathway and a caudal 'spatial' pathway was not apparent when simple tones (designed to be analogous to simple visual features) were used as stimuli[8]. It was only when the vocal calls of monkey conspecifics were used that these properties became obvious[9]. Furthermore, there is also strong evidence that different kinds of auditory

information are represented in distinct parts of the brain; for example, stroke can rob someone of the ability to understand music while preserving functions such as the comprehension of speech and other sounds[10]. Nevertheless, domain-specific approaches to understanding audition cannot (or do not aim to) account for the perception and processing of sounds outside these domains (such as impact sounds, which are neither vocal nor musical). What is therefore needed is a domain-general model in which there are multiple interacting computations, such as those that have been proposed for vision[11].

Recent developments in auditory neuroscience have begun to reveal candidate organizational principles for the processing of sound in the primate brain[12–14]. In this article, we argue that these organizational principles can be used to develop more computationally driven, domain-general models of cortical auditory processing. Previous reviews on auditory processing have characterized the involvement of rostral and caudal pathways with specific auditory and linguistic domains[1–7]. Other accounts have posited the relationship of these pathways to attention[15,16] or described

their role in perceiving auditory objects[17]. Our purpose here is rather different. We describe and synthesize recent findings of auditory neuroscience studies that have used neuroanatomical analyses, electrocorticography (ECoG) and functional MRI (fMRI) in humans and monkeys with the aim of setting out a domain-general functional account of the primate auditory cortex. The model that we propose is based on rostro–caudal patterns of intracortical and extracortical connectivity in the auditory cortex, the differential temporal response properties of rostral and caudal cortical fields and task-related functional engagement of the rostral and caudal regions of the auditory cortex.

Auditory anatomical organization

In audition, the signal carried by the auditory nerve is deconstructed into different kinds of informational features, which are represented in parallel in the ascending auditory pathway (BOX 1). Within these representations, some general organizational principles are apparent. Tonotopy — in which the frequency information in sound is represented across a spatial array — is first established in the cochlea and is preserved along the entire ascending auditory pathway[18]. In addition, other acoustic features — such as sound onsets and offsets, temporal regularities relating to pitch, and spatial location — are computed from the cochlear nucleus onwards[18]. Thus, there is intense complexity in the subcortical processing of sound, and this complexity (BOX 2) is preserved even as the temporal detail of the sound representations decreases (BOX 1). Following this subcortical processing, the medial geniculate body (auditory thalamus) projects to the cortex (which also makes strong connections back to subcortical nuclei; FIG. 1a).

The primate auditory cortex is organized, anatomically, in a rostral–caudal orientation, with three core primary fields surrounded by belt and parabelt fields, in a roughly concentric form. The tonotopic organization seen in the ascending auditory pathway is seen within the core fields, with three different tonotopic gradients seen across the three core fields[8,19]. Connectivity within the 'core' auditory cortex also maintains a rostral–caudal axis, with greater connectivity

## Box 1 | The ascending and descending auditory pathways

Before sound is represented in the auditory cortex, it is first decomposed and undergoes extensive analysis in the ascending auditory pathway (see the figure). For example, the spatial properties of sounds are known to be computed subcortically[8,9,96]; thus, it is assumed that they do not need to be re-computed cortically. This subcortical processing is supplemented by further processing through cortico-thalamic loops to enable auditory perception.

At the cochlea, the physical vibrations that give rise to the perception of sound are transduced into electrical signals. The cochlea encodes sound in a tonotopic form; that is, sounds of different frequencies are differentially represented. This tonotopic information is preserved within the auditory nerve and throughout the entire ascending auditory pathway into the core auditory cortical fields[18]. The auditory nerve fibres project from the cochlea to the cochlear nucleus (see the figure), where the auditory signal is decomposed into a number of parallel representations[18]. Divided into dorsal, anteroventral and posteroventral portions, the cochlear nucleus contains six principal cell types (as well as small cell and granule cell types) and mediates immensely complex processing of the auditory signal, which is only roughly characterized here. Each population of particular cochlear nucleus cell types receives input from across the whole tonotopic range and projects to a specific set of brainstem fields[97]. The anteroventral cochlear nucleus (AVCN) contains cells that respond to sounds with a very high level of temporal precision[18]. These project principally to the superior olivary nucleus and the trapezoidal body, which are important in computing the spatial location of sounds by comparing the inputs from the two ears, and thence to the inferior colliculus (IC)[18]. The posterodorsal cochlear nucleus (PVCN) contains cells that respond to sound onsets and exhibit repeated regular ('chopping') responses to sustained sounds. These PVCN cells display a broader range of frequency responses than those in the AVCN[18]. The dorsal cochlear nucleus (DCN) contains cells that display very complex frequency responses, such as highly specific frequency combination responses[18]. This may enable the identification of spectral 'notches', which are energy minima in the distribution of sound energy over frequency that are important for perceiving the spatial location of sound in the vertical plane. In addition to projecting to the superior olivary nucleus and trapezoidal body, the AVCN and PVCN both directly project to the lateral lemniscus and the IC[18]. The cochlear nucleus thus contributes to different sound processing pathways and contributes to the detection of a wide range of different informational aspects of incoming sounds, such as the spatial location of the source of the sound or the properties of the sound that can contribute to its identification (such as its pitch)[97].

Further along the pathway, the IC is a critical relay station in the processing of sound. Tonotopy is preserved and neurons are organized in sheets of cells that share common frequency responses. However, within a sheet, neurons can vary in their responses to other aspects of sounds, such as their spatial location and amplitude characteristics[97]. Neural representations in the IC are less affected by noisy and reverberant auditory environments than those of cochlear nucleus neurons, suggesting that the processing between these two regions makes the signal more robust, which may aid consistency in perceptual experience[98].

The IC projects to the auditory thalamus (including the medial geniculate nucleus, the medial pulvinar (PM) and the suprageniculate (Sg)–limitans (Lim) complex). The ventral medial geniculate nucleus (MGv)

is, similar to the IC, organized tonotopically and is considered to be the main pathway to the auditory cortex, although other thalamic nuclei project to auditory fields (FIG. 1a). The medial geniculate nucleus (MGm) receives auditory, visual somatosensory and vestibular inputs, and dorsal geniculate nuclei (MGd) also receive auditory and somatosensory inputs; these cells tend to have fast, frequency-specific responses to sounds[97]. These thalamic nuclei project to the auditory core and surrounding auditory fields in the cortex[13] (FIG. 1a).

It is important to note that the primate auditory system does not faithfully transmit the auditory environment to the cortex. There is considerable loss of spectral detail at the cochlea, with a roughly logarithmic relationship between frequency and resolution, meaning that the higher the frequency of the sound, the more compressed its resolution[99]. However, there is reasonably good resolution of temporal detail at the cochlea, which is essential for the encoding of the interaural time differences that are used to compute spatial location of sounds[100]. At the IC, amplitude modulations with modulation rates slower than 200–300 Hz (that is, those with a repetition rate of approximately 3.3 ms and longer) can be processed. However, this temporal sensitivity reduces as the sounds are processed in the ascending auditory pathway[101]. For this reason, humans are perceptually poor at detecting amplitude modulations with modulation rates that are faster than 50–60 Hz (that is, those with a repetition length of 16–20 ms or longer)[102]. DNLL, dorsal nucleus of the lateral lemniscus; ICc, central nucleus of the IC; ICdc, dorsal cortex of the IC; ICp, pericentral nucleus of the IC; ICx, external nucleus of the IC; LSO, lateral superior olivary nucleus; MNTB, medial nucleus of the trapezoid body; MSO, medial superior olivary nucleus; VNLL, ventral nucleus of the lateral lemniscus.

Decreasing temporal resolution
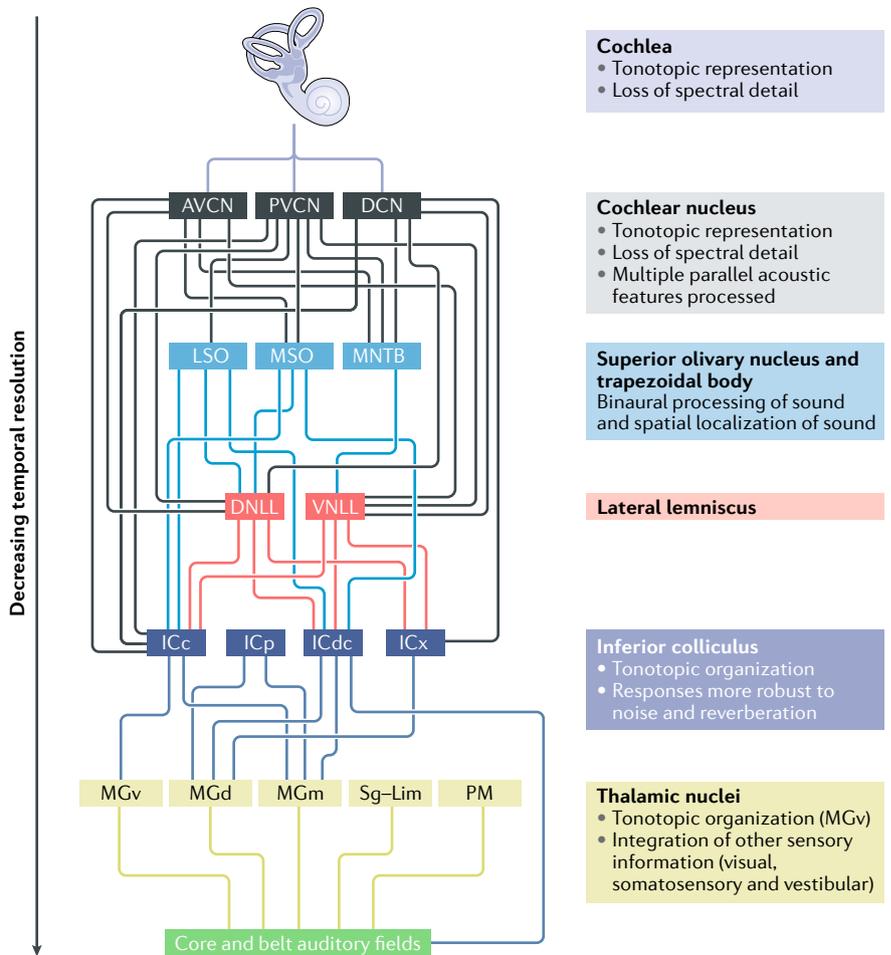
Cochlea
• Tonotopic representation
• Loss of spectral detail

AVCN PVCN DCN

Cochlear nucleus
• Tonotopic representation
• Loss of spectral detail
• Multiple parallel acoustic features processed

LSO MSO MNTB

Superior olivary nucleus and trapezoidal body
Binaural processing of sound and spatial localization of sound

DNLL VNLL

Lateral lemniscus

ICc ICp ICdc ICx

Inferior colliculus
• Tonotopic organization
• Responses more robust to noise and reverberation

MGv MGd MGm Sg–Lim PM

Thalamic nuclei
• Tonotopic organization (MGv)
• Integration of other sensory information (visual, somatosensory and vestibular)

Core and belt auditory fields

Figure is adapted with permission from REF.[19], Proceedings of the National Academy of Sciences.

Although both visual and auditory perceptual pathways share similarities (without which cross-modal perceptual benefits would be impossible), there are a number of important differences between auditory and visual processing in terms of anatomy and computational constraints. For example, although the number of synaptic projections in the ascending visual and auditory pathways is similar, there are more synaptic connections in the retina, with more cell types and more complex connectivity[103] than there are in the cochlea[18]. By contrast, there are more nuclei involved in the subcortical processing of sound than there are in the visual pathway, with a great deal of decomposition of the auditory environment and auditory objects taking place in the ascending auditory pathway (BOX 1). As a result, visual perception relies heavily on cortical processing, arguably more so than audition does[98]. Indeed, damage to the primary visual cortex (V1) causes cortical blindness — a loss of the visual field that cannot be recoded or recovered. Thus, patients with V1 damage cannot report on visual information presented to the corresponding parts of the visual field[104]. However, bilateral damage to the primary auditory cortex does not lead to cortical deafness — sounds can still be heard, but the processing of structural information in the sound (which is required to recognize speech) is not possible[105]. Such patients are thus typically described as being 'word deaf'. Similarly, V1 represents a map of the input to the retina, whereas primary auditory fields show a less invariant response and have been argued to show a more context-sensitive profile; that is, different neural responses are generated in the primary auditory cortex to the same sound, depending on the frequency with which it is presented[106]. This may suggest that auditory perception is more heterogeneous and flexible than visual perception, perhaps enabling animals to deal with considerable variation in auditory environments[107].

Unlike the visual system, in which spatial information is encoded as part of the representation at the retina and V1, auditory spatial information is computed (largely) by making comparisons across the two ears, and this occurs from early stages of the ascending auditory pathway[108]. This contributes to the construction of representations of the auditory objects in our environment. These representations can be based on low-level computations, such as spatial location, spectral shape and sequential information, or on higher-order knowledge and can entail cross-modal processing (seen in the 'ventriloquist effect', for example)[109].

Unlike visual objects, sounds only exist in our environment because something has happened. That is, sounds are always caused by actions, and when sounds are produced, we hear the properties both of the objects that the sound was made with and the kinds of actions that were made with them. For example, hands make a different sound when they are clapped together than when they are rubbed together, a stringed musical instrument will make a different sound when it is plucked or when it is bowed and a larger stringed instrument will produce sounds of a different pitch and spectral range than a smaller one, no matter how it is played. By contrast, many visual objects merely require visible light to be reflected from them for us to be able to perceive them, which is even true for moving visual objects (which of course also have a structure that evolves over time, similar to sound).

The strong link between sounds, objects and actions may also underlie the robust finding that auditory sequences are far better than visual sequences for conveying a sense of rhythm[110]. The link between sound objects and actions also means that sounds can convey a great deal of information without necessarily being specifically recognized. A loud impact sound behind me will cause me to react, even if I cannot recognize exactly what hit what — it suggests that something large hit something else hard, and whatever hit what, I might want to get out of the way.

between adjacent core auditory regions than between non-adjacent core fields[12] (FIG. 1b). This rostral–caudal organization is also seen in the connections between the auditory thalamus and the rostral and caudal core auditory fields; caudal core field A1 and the rostral auditory core field (R) both receive the vast majority of their inputs from the ventral medial geniculate body, whereas the rostral-most core field, the rostral temporal core field (RT), receives a greater proportion of inputs from the posterodorsal medial geniculate body[13] (FIG. 1a). Rostral–caudal differences extend into the thalamo-cortical connectivity of the rostral belt and parabelt areas. The rostrotemporal polar belt area (RTp), lying directly rostral to RT, receives most of its inputs from the posterodorsal,

ventral and medial fields within the medial geniculate body, whereas the rostral superior temporal gyrus (STGr), lying lateral to RTp in the parabelt, is more strongly connected to the medial pulvinar and suprageniculate (Sg)–limitans (Lim) complex[13].

This rostral–caudal organization of anatomy and connectivity has been taken as contributing evidence to support the idea that the nature of processing in the rostral and caudal auditory cortex may be qualitatively distinct. For instance, it has been suggested that rostral projections and fields may be more likely (than caudal projections) to play a fundamental role in the integration of audiovisual information because they are more strongly anatomically connected to thalamic nuclei that process
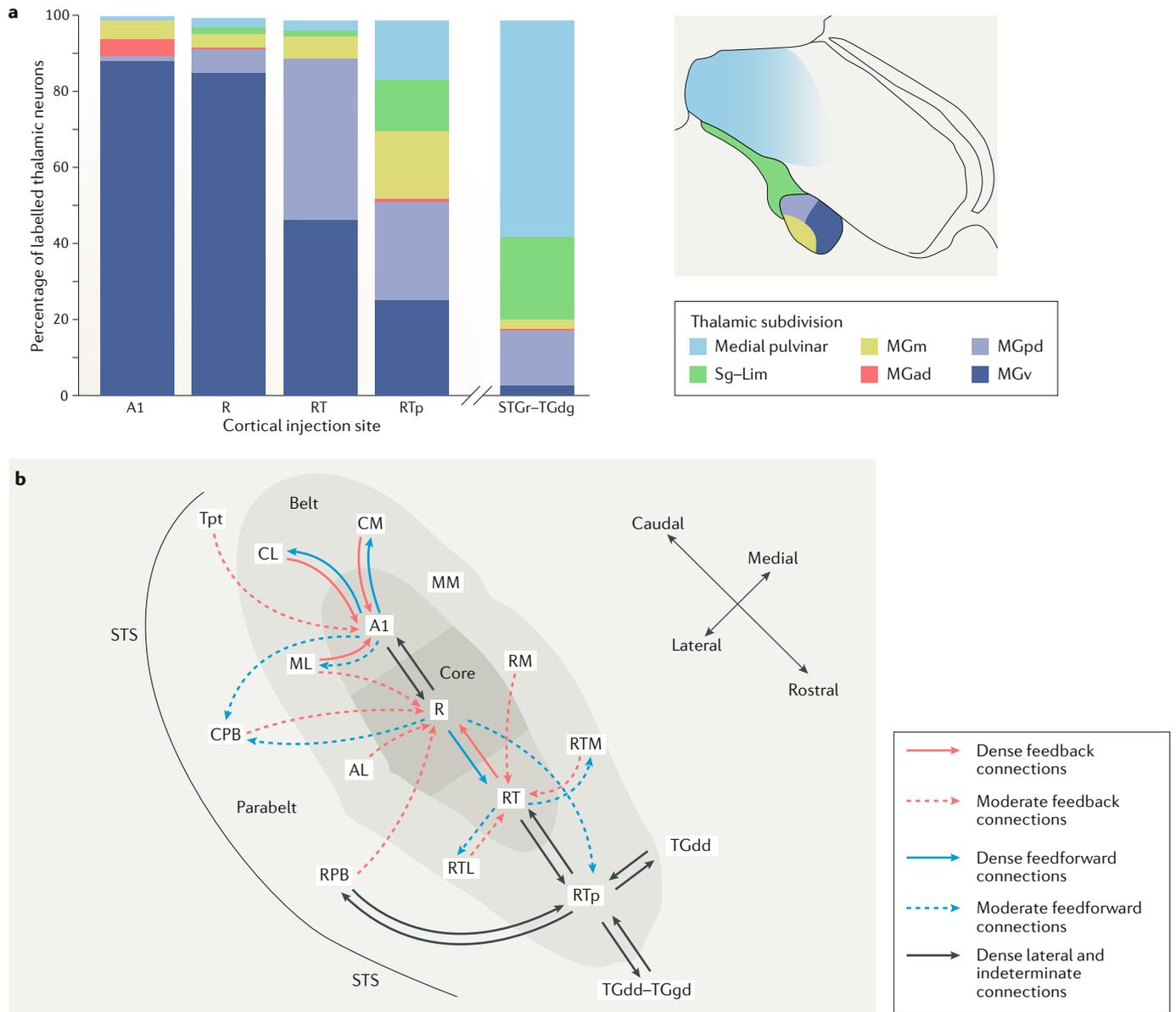
visual information as well as sound[13] (the medial pulvinar and Sg and Lim thalamic nuclei; FIG. 1a). Caudal areas, on the other hand, are proposed to be involved with processing audio-somatosensory stimuli, responding both to sounds (such as clicks) and to facial somatosensory stimulation[20,21], and may mediate the roles of facial somato-sensation and sound processing in the control of articulation[22]. In support of this proposal, caudal belt regions do not receive inputs from the visual thalamus but do show (in addition to auditory thalamus connectivity) input from the somatosensory thalamus[21].

Auditory response properties

The possibility of differences in the perceptual processing properties of rostral and caudal areas of the auditory cortex suggested by their anatomy and connectivity is supported by differences in the response properties exhibited by neurons in these regions, as described below. However, it is first worth noting the many similarities between these areas. In terms of representing the frequency of sound, the tonotopy encoded at the cochlea is preserved in each of the core auditory fields (although it reverses directions across fields at the boundaries of core auditory areas[11,19,23]). Neurons in the rostral and caudal auditory cortex are also similar in their frequency tuning (the breadth of the range of frequencies that each cell responds to), their response threshold (how loud a sound has to be to stimulate the cell) and their activation strength (the average driven spike rate)[14].

Nevertheless, important rostral–caudal differences can be seen in the speeds of neural responses and neural sensitivities to the structure of sounds over time[24]. There are rostral–caudal differences in response latency in both core and belt areas[25]. Caudal core area A1 shows a faster median response to sounds (within 20 ms of onset) than the more rostral area R (33 ms)[14] (FIG. 2), and caudomedial belt areas have been shown to have an average response latency of approximately 10 ms, even faster than core areas[26]. Area A1 also tracks fast acoustic amplitude modulations (with a duty cycle on the order of 20–30 ms) more accurately than the more rostral core area R, which can track only slower amplitude modulations (with a duty cycle on the order of 100 ms and above)[14]. Neurons in area R saturate in their response at lower frequencies than do those in A1 (FIG. 2), indicating that neurons in area R lose synchrony at lower rates than those in A1, which can continue to synchronize to faster rates of amplitude modulation[14].

Fig. 1 | **Cortical and subcortical connectivity of the macaque auditory cortex. a** | Distribution of inputs to each auditory cortical field from the different thalamic nuclei. There is a clear rostral–caudal shift in thalamic connectivity. Moving rostrally, there is a general decline in the proportions of connections from the ventral division of the medial geniculate nucleus (MGN) of the thalamus (MGv) and increased proportions of inputs from other MGN and other thalamic nuclei. Core areas A1 and R receive an overwhelming majority of their inputs from the MGv, whereas the more rostral temporal core field (RT) area receives similar proportions of inputs from the MGv and the posterodorsal MGN (MGpd). The rostrotemporal polar belt area (RTp) receives roughly similar proportions of its inputs from the MGv, MGpd and the medial division of the MGN (MGm) as well as the suprageniculate (Sg)–limitans (Lim) complex and the medial pulvinar. The rostral superior temporal gyrus (STGr) and the granular part of the dorsal temporal pole (TGdg) receive the majority of their thalamic inputs from the medial pulvinar and a lower proportion from the Sg–Lim complex and the MGpd[13]. **b** | A schematic image illustrating the connectivity of different core auditory regions in the macaque cortex[12,19,111]. Dense connections (those for which the retrograde connectivity cell count was over 30) are represented with solid lines, whereas moderate connections (those for which cell count was between 15 and 29) are shown with dashed lines.

The connectivity pattern shows a clear rostral–caudal difference. Caudal core field A1 primarily connects to surrounding belt fields and to the rostral auditory core field (R), with more moderate connections to the caudal belt and parabelt fields. R, on the other hand, connects to A1 and to the RT, with moderate connections to the rostral and caudal belt and parabelt fields and the RTp. RT connects to the adjacent field RTp and to the adjacent rostral belt fields. RTp has a distinctly different pattern of connectivity from the temporal pole, rostral belt and parabelt fields via lateral and indeterminate connections. This pattern of connectivity results in a recurrent and interactive network incorporating multiple parallel pathways with both direct and indirect connections[12]. AL, anterolateral belt; CL, caudolateral belt; CM, caudomedial belt; CPB, caudal parabelt; MGad, anterodorsal division of the MGN; ML, middle lateral belt; MM, middle medial belt; RM, rostromedial belt; RTL, rostrotemporal-lateral belt; RTM, rostrotemporal-medial belt; RPB, rostral parabelt; STS, superior temporal sulcus; TGdd, dysgranular part of the dorsal temporal pole; Tpt, temporo-parietal area. Part **a** is adapted with permission from REF.[13], Wiley-VCH. Part **b** is adapted with permission from Scott, B. H. et al. Intrinsic connections of the core auditory cortical regions and rostral supratemporal plane in the macaque monkey. *Cereb. Cortex* (2017) **27**(1): 809–840, by permission of Oxford University Press (REF.[12]).
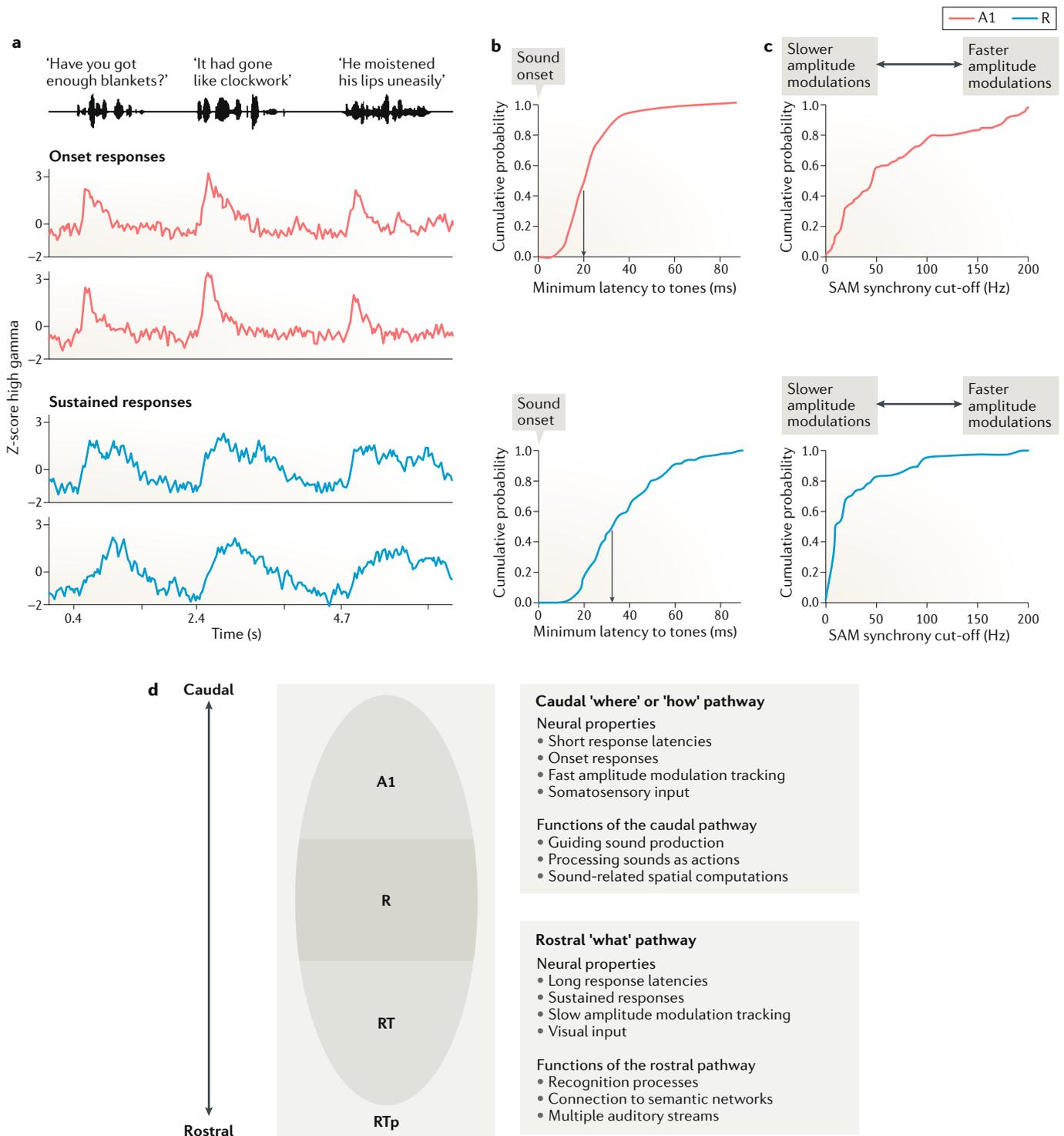
Fig. 2 | **Response properties of the rostral and caudal auditory cortex.**
a | Examples of electrocorticography responses to sentences that have been categorized as 'sustained' and 'onset' responses on the basis of machine learning classifications. Caudal fields (red traces) show transient responses associated with the onset of complex sequences and rostral fields (blue traces) show sustained responses. These distinctions were found bilaterally[31]. **b** | Minimum response latencies (that is, the fastest responses to sound onsets) in caudal core field A1 (top, red) and rostral core field R (bottom, blue). The median response in caudal A1 is faster (at 20 ms) than that in rostral R (33 ms)[14]. **c** | Neural responses to increasing rates of sinusoidal amplitude modulation (SAM) in caudal core field A1 (top) and rostral core field R (bottom). The SAM synchrony cut-off reflects the amplitude modulation rate at which a neuron's synchronization to the amplitude envelope can no longer be detected. Note that the responses saturate at a much lower amplitude modulation frequency in rostral field R (median 10 Hz) than in caudal field A1 (median 46 Hz), indicating that the responses in A1 can track amplitude changes at a much faster rate than can R[14]. **d** | Schematic of the auditory cortex showing candidate functional distinctions in rostral and caudal areas. RT, rostral temporal core field; RTp, rostro-temporal polar belt area. Part **a** is adapted with permission from REF.[31], Elsevier. Parts **b** and **c** are adapted with permission from REF.[14], American Physiological Society.

In other words, the caudal core auditory cortex responds quickly to sound onsets and tracks fast amplitude modulations accurately, whereas the rostral auditory cortex responds more slowly to the starts of sounds and more accurately tracks slower amplitude modulations (such as those found in speech).

Although these rostral–caudal temporal processing differences are not completely distinct, and the similarities in response properties of core areas should not be ignored[27], we hypothesize that they may relate to important functional differences at higher levels of the auditory cortex. The faster and more precise temporal response in caudal A1 suggests that caudal auditory fields may more accurately compute certain aspects of sound sequences than more rostral fields. For example, the perception of rhythm is based on perceived beat onsets in sounds. The finer temporal acuity of caudal areas may make them better at tracking and coding these perceived onsets than the rostral auditory cortex, which also has a poorer resolution of different amplitude modulation frequencies (perceived beat onsets have been linked to amplitude onsets as opposed to amplitude offsets[28]). We hypothesize that this difference in temporal acuity may also make the caudal auditory cortex suitable for performing computations that guide actions, which need to occur quickly if they are to be of any utility in the control of movement. There is evidence that engaging the motor system in an auditory perception task does indeed increase the temporal acuity of responses, which may reflect the enhanced involvement of caudal auditory fields. For instance, it is known that participants are more accurate at tracking changes in auditory intervals when they are tapping along than when they are passively listening, and we would suggest that this reflects differential recruitment of caudal auditory sensory motor systems, which are recruited by coordinating actions with sounds[29].

In contrast to the brisk onset responses seen in caudal fields, we suggest that the slow onset times seen in rostral fields may reflect processes that are slower and that entail feedforward and feedback patterns of connectivity. Circuits mediating hierarchical perceptual processing and recognition processes, for example, tend to be slower in their responses, which reflects the time courses of prediction and integration of incoming perceptual information with prior experience (for example, the use of context in understanding)[30]. Indeed, auditory recognition processes can be relatively slow. In humans, electrophysiological studies show that the earliest neural correlates of

auditory semantic processing can be seen about 200 ms after stimulus presentation and continue to unfold over a further several hundred milliseconds[30]. We therefore suggest that rostral auditory cortex areas may be well suited to a role in such processes.

Further evidence for the distinct temporal response properties of the rostral and caudal auditory cortex comes from human studies. In a recent ECoG experiment[31], cortical recordings were obtained as 27 participants heard spoken sentences. An unsupervised learning approach clustered the neural responses according to their temporal profiles. The results revealed large rostral–caudal differences; caudal fields responded quickly and strongly to the onsets of sentences (with additional onset responses occurring after gaps in speech longer than 200 ms). Rostral fields, by contrast, showed much weaker onset responses and produced slower and more sustained responses. This difference (FIG. 2) supports the idea that there are computational differences between rostral and caudal auditory cortical fields in terms of basic acoustic processing, with caudal fields being more sensitive to onsets (and hence the temporal characteristics of sounds) and rostral fields being more sensitive to the spectro-temporal information conveyed over the whole sequence of a sound. Indeed, pure tones, which do not contain changing structure over time, produced only caudal onset responses. The results were seen over all stimuli (including nearly 500 natural sentences and single syllables) and did not depend on the linguistic properties of the sentence or the phonetic properties of the speech sounds, indicating that this may represent a more global rostral–caudal distinction in temporal response characteristics. We note, however, that ECoG study participants are almost always patients with intractable epilepsy who are on medication and may have suffered brain tissue damage or trauma. Thus, the results from such studies should be interpreted cautiously and corroborated with evidence obtained with other techniques.

Human functional imaging has revealed processing differences in rostral and caudal auditory areas that are in line with the monkey electrophysiology and human ECoG findings discussed above. In one study, participants were presented with a variety of different kinds of sounds (including speech, emotional vocalizations, animal cries, musical instruments and tool sounds). Their cortical responses (measured with fMRI) were analysed with respect to the

spectro-temporal features of the sounds[32]. The presentation of sounds in which there were fast modulations of the amplitude envelope (that is, the changes in the amplitude of a sound over time) but slower spectral modulations (that is, changes in the large-scale distribution of the frequency content of the sounds, such as those that characterize formants in speech) led to an enhanced response in medial regions caudal to Heschl's gyrus (the major anatomical landmark for primary auditory fields in the human brain), whereas sounds that contained more detailed spectral information and broader amplitude envelope modulations were associated with responses in regions rostral to Heschl's gyrus and in the STGr. This pattern was replicated in a second fMRI study that examined responses to environmental sounds, speech and music[33]. Caudal auditory fields responded preferentially to fast amplitude envelope modulations and slower spectral modulations, whereas rostral fields responded preferentially to faster spectral modulations and slow amplitude envelope modulations. As in the experiments in monkeys, these response profiles suggest that caudal and rostral regions may be involved in distinct computations; rostral fields may process information conveyed in the spectral detail of a sound, whereas caudal fields may process information conveyed via the amplitude envelope. Below, we examine the more specific functional properties of rostral and caudal auditory regions that these neuronal differences may indicate.

Rostral auditory processing
Recognition processes. Human speech is a perfect example of a spectrally complex sound. Comprehending speech requires the auditory system to grapple with dynamic changes in the spectral profile and, from this, recognize meaningful units of sound (such as phonemes, words and grammatical and prosodic structures). Given the rostral auditory cortex's proposed role in processing spectrally complex sounds, it is unsurprising that activity levels in rostral auditory fields are highly sensitive to the intelligibility of speech[34] and that subfields within the rostral STG respond to specific components of intelligible speech such as syntactic structures[35,36].

On the basis of a review of the literature, it has been argued that there is a caudal–rostral hierarchical processing gradient for speech (mainly in the left cortical hemisphere) in which the most basic acoustic processing takes place in the primary auditory cortex (PAC) and the complexity of processing increases as the information progresses rostrally,

with the processing of higher-order lexical and semantic structure taking place near the temporal pole[37]. This proposed hierarchy mirrors the gradient in cortical thickness measurements in temporal areas; the cortex is thinner and there are fewer feedback connections crossing cortical layers near the PAC, whereas the cortex is thicker and has a higher ratio of feedback connections (those from deeper cortical layers to superficial cortical layers) to feedforward connections near the temporal pole[38]. The greater number of connections across layers has typically been assumed to be linked to greater processing complexity[38]. Furthermore, physiological studies in non-human primates have shown that rostral STG auditory areas exhibit more inhibitory responses than excitatory responses and that the latencies of these responses are longer than they are in more caudal areas — properties indicative of a higher position in a hierarchical processing stream[39]. It is also the case that rostral superior temporal lobe responses to speech appear to be malleable and sensitive to the effects of prediction and context. Whereas mid-STG fields are unaffected by sentence expectations, rostral auditory areas respond selectively on the basis of the expected or violated sentence endings[40]. Notably, this more context-sensitive response is mediated by input from the larger language network and is associated with specific connectivity between rostral auditory areas and ventral frontal cortical fields[40,41].

Speech perception is perhaps the most well-studied auditory recognition process; however, the processing of other sorts of spectrally complex auditory objects (such as birdsong or instrumental music) also recruits the rostral auditory cortex[42]. Response biases in rostral auditory fields to particular sound classes, including speech, voices and music, can be detected using fMRI, but these effects are weak in that they are not purely selective (that is, auditory areas that respond to music also respond to other types of sounds)[33]. In addition, it has been noted that although a single study investigating a particular sound class may show a hierarchical response profile in which the responses become more selective along the rostral pathway, this does not imply that the rostral pathway as a whole is selective for that sound class[43].

Parallel processing of multiple auditory objects. In normal environments, we frequently hear multiple auditory objects simultaneously (at the time of writing,

for example, we can hear a car alarm and footsteps in the corridor in addition to the sounds of our own typing). We know that unattended auditory information can disrupt performance in behavioural tasks requiring speech production or holding verbal information in working memory, which suggests that unattended auditory objects are being processed (to some extent) for meaning in parallel with attended auditory information[44]. The ascending auditory pathways are essential for forming representations of auditory objects, and their associated spatial locations (BOX 1) and rostral cortical auditory fields appear to be capable of representing multiple parallel auditory objects, only one of which forms the currently attended signal[45,46]. Studies of 'masked' speech, in which a target speech signal is heard against a simultaneous competing sound, indicate that when a competing sound is more speech-like, it elicits a greater neural response in rostral auditory fields[47,48]. This response occurs in addition to the activation associated with the content of the attended speech[47] (which may include self-produced speech[48]), suggesting that the computational processes taking place in the rostral auditory cortex must be flexible enough to process (and recognize aspects of) multiple unattended auditory objects. This flexibility must permit the processing of multiple parallel sources of auditory information for a wide variety of possible kinds of sound as well as the switching of attentional focus between them. Such switching may occur on the basis of intention and/or when information in an unattended stream starts to compete for resources with the attended stream[49]. Such parallel processing must therefore be fast, plastic and highly state-dependent.

Caudal auditory processing
Sensorimotor and spatial computations.
As discussed above, caudal auditory fields show precise and rapid responses to sound onsets and fluctuations. We suggest that this makes them ideal for guiding motor responses to sounds in the environment or to self-produced sounds, especially those that require rapid action. Speech production is, of course, a motor action that requires tight temporal and spatial control[50]. Caudal auditory cortical fields have been shown many times to be recruited during speech production[51–55], whereas the activity of rostral auditory fields is suppressed during articulation (relative to its activity when hearing speech)[56,57]. This motor-related caudal auditory activity is enhanced when a talker, for example, alters their voice to

match specific pitches[58], compensates for an experimentally induced shift in their perceived vocal pitch[59] or speaks (usually with considerable disfluencies) while being presented with delayed auditory feedback[60]. Superior auditory-motor abilities have been shown to correlate with neural measures in pathways connected to caudal auditory fields. For example, the arcuate fasciculus, a white matter tract that projects from the caudal temporal and parietal cortex to the frontal lobe, shows greater leftward lateralization in terms of volume and increased integrity (measured with fractional anisotropy) in people who are better at repeating words from foreign languages[61]. Conversely, difficulties with speech production (such as stammering) have been linked to abnormalities in pathways connected to caudal auditory fields[62,63].

Rostral auditory streams support recognition processes under normal listening conditions (see above); however, caudal areas do seem to play a limited role in recognition processes. Caudal areas are recruited only during some specific kinds of perceptual tasks, including those requiring sublexical units (such as phonemes)[64–67] and phonetic features[68] of speech to be accessed, motor-related semantic features to be processed (as is the case for Japanese onomatopoetic words[69]) or the passive perception of non-speech mouth sounds that 'sound do-able' (can be matched to an action)[22,70]. It is also important to note that when auditory recognition processes require an emphasis on the way that a sound is made (for example, when beat boxers hear unfamiliar examples of expert beat boxing[71] or people listen to sounds produced by human actions[72,73], such as the sounds made by hand tools[74]), caudal auditory areas are recruited and form part of a wider sensorimotor network.

Although speech is usually considered the prototypical sound-related action, the audiomotor integration of other types of sounds also relies on caudal fields. Musical performance, for example, requires precise cortical responses to sound in order to guide accurate motor production[75]. Although there is much less published research on the neuroscience of music than there is on speech, effects similar to those observed for speech (such as an enhanced caudal auditory cortex response) are seen when study participants attempt to perform music while receiving perceptual feedback that is altered[76]. Action-related sounds also guide many other movements (for example, the rising resonance frequency as a glass of water is filled indicates when

to stop pouring). Similarly, the sound an egg makes when it cracks, and many other action-related sounds, require precise responses[77,78], and startle reflexes to loud sounds entail an immediate orientation to the perceived sound location. Although we know of no published studies on motor guidance in response to such sounds, we predict that they should also recruit caudal auditory cortex fields.

Movement is, of course, closely linked with space. The bias for responses to the onsets of sounds in the caudal auditory cortex[31], combined with the capacity of caudal areas to produce fast and fine temporal responses to sounds, could make these areas suitable for representing the spatial locations of sounds[79] and guiding and processing movement[80] and navigation (as described in the following studies) accordingly. Recent evidence from fMRI studies that have used binaural and 3D sound presentation paradigms supports this. Blind human participants showed increased caudal temporal (and parietal) responses to echoes when listening to recordings that were binaural (and therefore contained the information necessary for echolocation) compared with when the recordings were monaural[81]. Similarly, in sighted people, sounds presented binaurally to create the illusion of a source existing outside the head activate caudal pathways more strongly than those presented monoaurally (which lack the information necessary to calculate a spatial origin and therefore appear to originate inside the head[82]). Caudal superior temporal activity is also modulated by varying the perceived location of a sound in space, as indicated by its direction[82] and proximity to the head[83]. However, single-cell recording studies in caudal belt fields find partially segregated responses to temporal features and spatial location, which suggests that two independent streams may contribute to these sensorimotor and spatial processes[26].

fMRI is an inherently correlational technique, but these computational distinctions are also supported by causal evidence obtained from a transcranial magnetic stimulation (TMS) study and a patient study. Whereas transient TMS applied to the rostral auditory cortex delayed reaction times for judgements concerning sound identity more than it affected those related to sound location, similarly disrupting the caudal auditory cortex delayed judgements of sound location more than it did sound identity[84]. Similarly, stroke damage to rostral areas affects sound identification, whereas damage to the caudal auditory cortex impairs location judgements[85].

In more ecologically valid contexts in which individuals are moving and talking with other people in complex auditory environments, one can imagine that sound identification processes and spatiomotor processes must interact. Indeed, in the processing of multiple auditory sources, spatial information about a sound is a powerful way of separating it out from other competing sounds. Although this separation likely has its origin in subcortical processes, caudal auditory fields may also be involved in aspects of the spatial representations of the sounds[86,87]. Similarly, recognizing a sound may be important for selecting the correct response action. It is likely that this interaction involves integration in frontal cortical areas, where the functional auditory pathways are proposed to converge[1].

Future directions

There is much still to uncover. It is very unlikely that neural temporal response differences to sound are the only relevant computational factors distinguishing the rostral and caudal auditory streams. Sounds are made by objects and actions, and the exact amplitude and spectral profile of any sound will reflect these underlying objects and actions in a complex way. Thus, the spectral and amplitude envelope properties of a sound will not be easily separated into those concerning identification and those related to spatial and motor functions. The computational processing of sounds for different purposes likely entails differential aspects of these amplitude and spectral characteristics. For example, caudal fields may be more sensitive to the amplitude onset of a sound, whereas rostral fields may be sensitive to the amplitude envelope of the whole sound.

An important question concerns what exactly the PAC represents given that so much structural information, including spatial location, is computed and coded in the ascending auditory pathway (BOX 1). Perhaps its role is to represent sound in such a way that it can be accessed by higher-order perceptual and cognitive systems. Indeed, the PAC has been shown to be highly non-selective to particular sound types (exhibiting no selectively greater response to speech sounds than other sounds, for example[88]) but conversely to be acutely sensitive to the context in which a sound is occurring. For example, it shows repetition suppression (an attenuated neural response to repeated stimuli)[89].

We also still do not know exactly how multiple auditory objects are represented,

processed and selected between in rostral auditory fields or precisely what kinds of auditory information are used to guide action. Is it really the case that the fine temporal sensitivity of caudal fields is matched by a weaker reliance on spectral cues[90] or is the system more complex than this? When we understand what aspects of sounds are represented at distinct levels of both cortical and subcortical processing, the corresponding acoustic profiles and the resulting functional responses (that is, how they are used), we will have moved closer to a computational model of primate hearing.

Several previous papers have put forward models of the properties of separate auditory processing streams[1–3,5,7,90,91]. Our proposal is distinct in that we are trying to synthesize across a wider range of auditory domains than the previous domain-specific models, and we have taken temporal response properties of neurons to sound as a feature to distinguish the two candidate systems' computational differences. A couple of previous approaches have used temporal processing as a way of distinguishing differences in auditory processing; however, both focused on the temporal characteristics of sounds and used this as a way of hypothesizing candidate processing differences between the left and right auditory fields. One model[92] suggested that, by analogy with the construction of spectrograms, the left auditory fields had good temporal resolution and poor spectral resolution, whereas the analogous regions on the right had poor temporal resolution. Another[93] specifically suggested that the left auditory fields sampled sounds at a faster rate than the right auditory fields, with a general model of 'window size' being shorter in the left and longer on the right. Both of these approaches aim to account for hemispheric asymmetries in speech and sound processing by positing selective processing of particular acoustic characteristics. However, we believe that the evidence for such specificity of acoustic processing is sparse[50]. Other previous reviews and meta-analyses have focused only on the functions of auditory pathways (that is, how they interact with factors such as attention[15,16] or their roles in segmenting continuous sound into discrete auditory objects)[17]. What we are suggesting in this article, which diverges from previous accounts, is that the temporal response characteristics of the rostral and caudal auditory cortical fields are distinct and may underlie computational differences that give rise to previously observed functional differences.

Concluding remarks

There are well-established anatomical differences in the cortical and subcortical connectivity of core auditory fields, and these have been linked to differential processing characteristics associated with different kinds of perceptual tasks. In non-human primates, these hypotheses have usually come about on the basis of single-cell recording studies, whereas in humans evidence has been primarily provided by functional imaging. Here, we have argued that a key feature of these processing differences is the temporal response characteristics of subregions of the auditory cortex. Differences in the temporal response characteristics of the rostral and caudal auditory cortex have been reported in non-human primates over the past decade[25,26]. More recently, the rostral–caudal connectivity of the auditory cortex has been further elaborated[12,13], and we have begun to see different temporal response characteristics to sound in the human brain[31]. Perhaps because of the extreme salience of heard speech as a vehicle for linguistic and social communication, or perhaps because of the clear clinical need to understand aphasia, cognitive neuroscience has often approached the understanding of the auditory cortex in a manner that has been largely focused on spoken language[2]. This may have obscured more general auditory perceptual processes that are engaged by speech but also perhaps by other sounds. Early studies demonstrated a role of rostral auditory fields in the comprehension of speech and for caudal fields in the processing of the spatial location of sounds and auditory sensory guidance of speech production[51,56,94,95]. This role can now be extended to a more general model in which auditory recognition processes take place in rostral fields whereas caudal fields play a role in the sensory guidance of action and the alignment of action with sounds in space. We suggest that it is in the temporal response differences in rostral and caudal fields that the functional 'what' and 'where' and/or 'how' pathways originate.

Kyle Jasmin [ID] [1,2]*, *César F. Lima* [ID] [1,3] *and Sophie K. Scott* [ID] [1]*

[1]*Institute of Cognitive Neuroscience, University College London, London, UK.*

[2]*Department of Psychological Sciences, Birkbeck University of London, London, UK.*

[3]*Instituto Universitário de Lisboa (ISCTE-IUL), Lisbon, Portugal.*

*e-mail: kyle.jasmin.11@ucl.ac.uk; sophie.scott@ucl.ac.uk

1. Rauschecker, J. P. & Scott, S. K. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat. Neurosci.* **12**, 718–724 (2009).
2. Scott, S. K. & Johnsrude, I. S. The neuroanatomical and functional organization of speech perception. *Trends Neurosci.* **26**, 100–107 (2003).
3. Hickok, G. & Poeppel, D. Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition* **92**, 67–99 (2004).
4. Sammler, D. et al. Dorsal and ventral pathways for prosody. *Curr. Biol.* **25**, 3079–3085 (2015).
5. Schirmer, A. & Kotz, S. A. Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing. *Trends Cogn. Sci.* **10**, 24–30 (2006).
6. Zatorre, R. J., Chen, J. L. & Penhune, V. B. When the brain plays music: auditory–motor interactions in music perception and production. *Nat. Rev. Neurosci.* **8**, 547–558 (2007).
7. Alain, C. et al. "What" and 'where' in the human auditory system. *Proc. Natl Acad. Sci. USA* **98**, 12301–12306 (2001).
8. Rauschecker, J. P. Processing of complex sounds in the auditory cortex of cat, monkey, and man. *Acta Otoralyngol.* **532**(Suppl), 34–38 (1997).
9. Rauschecker, J. P. & Tian, B. Mechanisms and streams for processing of "what" and "where" in auditory cortex. *Proc. Natl Acad. Sci. USA* **97**, 11800–11806 (2000).
10. Rosemann, S. et al. Musical, visual and cognitive deficits after middle cerebral artery infarction. *eNeurologicalSci* **6**, 25–32 (2017).
11. Kravitz, D. J. et al. A new neural framework for visuospatial processing. *Nat. Rev. Neurosci.* **12**, 1–14 (2011).
12. Scott, B. H. et al. Intrinsic connections of the core auditory cortical regions and rostral supratemporal plane in the macaque monkey. *Cereb. Cortex* **7**, 809–840 (2015).
13. Scott, B. H. et al. Thalamic connections of the core auditory cortex and rostral supratemporal plane in the macaque monkey. *J. Comp. Neurol.* **525**, 3488–3513 (2017).
14. Scott, B. H., Malone, B. J. & Semple, M. N. Transformation of temporal processing across auditory cortex of awake macaques. *J. Neurophysiol.* **105**, 712–730 (2011).
15. Arnott, S. R. & Alain, C. The auditory dorsal pathway: orienting vision. *Neurosci. Biobehav. Rev.* **35**, 2162–2173 (2011).
16. Alho, K. et al. Stimulus-dependent activations and attention-related modulations in the auditory cortex: a meta-analysis of fMRI studies. *Hear. Res.* **307**, 29–41 (2014).
17. Bizley, J. K. & Cohen, Y. E. The what, where and how of auditory-object perception. *Nat. Rev. Neurosci.* **14**, 693–707 (2013).
18. Young, E. D. & Oertel, D. in *The Synaptic Organization of the Brain* (ed. Shepherd, G. M.) 125–164 (Oxford Univ. Press, 2004).
19. Kaas, J. H. & Hackett, T. A. Subdivisions of auditory cortex and processing streams in primates. *Proc. Natl Acad. Sci. USA* **97**, 11793–11799 (2000).
20. Smiley, J. F. et al. Multisensory convergence in auditory cortex, I. Cortical connections of the caudal superior temporal plane in macaque monkeys. *J. Comp. Neurol.* **502**, 894–923 (2007).
21. Hackett, T. A. et al. Multisensory convergence in auditory cortex, II. Thalamocortical connections of the caudal superior temporal plane. *J. Comp. Neurol.* **502**, 924–952 (2007).
22. Warren, J. E., Wise, R. J. S. & Warren, J. D. Sounds do-able: auditory–motor transformations and the posterior temporal plane. *Trends Neurosci.* **28**, 636–643 (2005).
23. Dick, F. et al. In vivo functional and myeloarchitectonic mapping of human primary auditory areas. *J. Neurosci.* **32**, 16095–16105 (2012).
24. Rauschecker, J. P. Where, when, and how: are they all sensorimotor? Towards a unified view of the dorsal pathway in vision and audition. *Cortex* **98**, 262–268 (2018).
25. Camalier, C. R. et al. Neural latencies across auditory cortex of macaque support a dorsal stream supramodal timing advantage in primates. *Proc. Natl Acad. Sci. USA* **109**, 18168–18173 (2012).
26. Kusmierek, P. & Rauschecker, J. P. Selectivity for space and time in early areas of the auditory dorsal stream in the rhesus monkey. *J. Neurophysiol.* **111**, 1671–1685 (2014).
27. Smith, E. H. Temporal processing in the auditory core: transformation or segregation? *J. Neurophysiol.* **106**, 2791–2793 (2011).
28. Scott, S. K. The point of P-centres. *Psychol. Res. Psychol. Forschung* **61**, 4–11 (1998).
29. Repp, B. H. & Keller, P. E. Adaptation to tempo changes in sensorimotor synchronization: effects of intention, attention, and awareness. *Q. J. Exp. Psychol. A* **57**, 499–521 (2004).
30. Holcomb, P. J. & Neville, H. J. Auditory and visual semantic priming in lexical decision: a comparison using event-related brain potentials. *Lang. Cogn. Process.* **5**, 281–312 (1990).
31. Hamilton, L. S., Edwards, E. & Chang, E. F. A spatial map of onset and sustained responses to speech in the human superior temporal gyrus. *Curr. Biol.* **28**, 1860–1871 (2018).
32. Santoro, R. et al. Encoding of natural sounds at multiple spectral and temporal resolutions in the human auditory cortex. *PLOS Comput. Biol.* **10**, e1003412–14 (2014).
33. Norman-Haignere, S., Kanwisher, N. G. & McDermott, J. H. Distinct cortical pathways for music and speech revealed by hypothesis-free voxel decomposition. *Neuron* **88**, 1281–1296 (2015).
34. Evans, S. et al. The pathways for intelligible speech: multivariate and univariate perspectives. *Cereb. Cortex* **24**, 2350–2361 (2014).
35. Agnew, Z. K. et al. Do sentences with unaccusative verbs involve syntactic movement? Evidence from neuroimaging. *Lang. Cogn. Neurosci.* **29**, 1035–1045 (2014).
36. de Heer, W. A. et al. The hierarchical cortical organization of human speech processing. *J. Neurosci.* **37**, 6539–6557 (2017).
37. Specht, K. Mapping a lateralization gradient within the ventral stream for auditory speech perception. *Front. Hum. Neurosci.* **7**, 629 (2013).
38. Wagstyl, K. et al. Cortical thickness gradients in structural hierarchies. *NeuroImage* **111**, 241–250 (2015).
39. Kikuchi, Y., Horwitz, B. & Mishkin, M. Hierarchical auditory processing directed rostrally along the monkey's supratemporal plane. *J. Neurosci.* **30**, 13021–13030 (2010).
40. Tuennerhoff, J. & Noppeney, U. When sentences live up to your expectations. *NeuroImage* **124**, 641–653 (2016).
41. Lyu, B. et al. Predictive brain mechanisms in sound-to-meaning mapping during speech processing. *J. Neurosci.* **36**, 10813–10822 (2016).
42. Leaver, A. M. & Rauschecker, J. P. Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. *J. Neurosci.* **30**, 7604–7612 (2010).
43. Price, C., Thierry, G. & Griffiths, T. Speech-specific auditory processing: where is it? *Trends Cogn. Sci.* **9**, 271–276 (2005).
44. Beaman, C. P. & Jones, D. M. Irrelevant sound disrupts order information in free recall as in serial recall. *Q. J. Exp. Psychol. A* **51**, 615–636 (1998).
45. Scott, S. K. Auditory processing — speech, space and auditory objects. *Curr. Opin. Neurobiol.* **15**, 197–201 (2005).
46. Zatorre, R. J. Sensitivity to auditory object features in human temporal neocortex. *J. Neurosci.* **24**, 3637–3642 (2004).
47. Evans, S. et al. Getting the cocktail party started: masking effects in speech perception. *J. Cogn. Neurosci.* **28**, 483–500 (2016).
48. Meekings, S. et al. Distinct neural systems recruited when speech production is modulated by different masking sounds. *J. Acoust. Soc. Am.* **140**, 8–19 (2016).
49. Brungart, D. S. et al. Informational and energetic masking effects in the perception of multiple simultaneous talkers. *J. Acoust. Soc. Am.* **110**, 2527–2538 (2001).
50. McGettigan, C. & Scott, S. K. Cortical asymmetries in speech perception: what's wrong, what"s right and what"s left? *Trends Cogn. Sci.* **16**, 269–276 (2012).
51. Hickok, G. A functional magnetic resonance imaging study of the role of left posterior superior temporal gyrus in speech production: implications for the explanation of conduction aphasia. *Neurosci. Lett.* **287**, 156–160 (2000).
52. Flinker, A. et al. Single-trial speech suppression of auditory cortex activity in humans. *J. Neurosci.* **30**, 16643–16650 (2010).
53. Agnew, Z. K. et al. Articulatory movements modulate auditory responses to speech. *NeuroImage* **73**, 191–199 (2013).

54. Jasmin, K. M. et al. Cohesion and joint speech: right hemisphere contributions to synchronized vocal production. *J. Neurosci.* **36**, 4669–4680 (2016).

55. Jasmin, K. et al. Overt social interaction and resting state in young adult males with autism: core and contextual neural features. *Brain* **142**, 808–822 (2019).

56. Wise, R. et al. Brain regions involved in articulation. *Lancet* **353**, 1057–1061 (1999).

57. Houde, J. F. et al. Modulation of the auditory cortex during speech: an MEG study. *J. Cogn. Neurosci.* **14**, 1125–1138 (2002).

58. Belyk, M. et al. The neural basis of vocal pitch imitation in humans. *J. Cogn. Neurosci.* **28**, 621–635 (2016).

59. Behroozmand, R. et al. Sensory–motor networks involved in speech production and motor control: an fMRI study. *NeuroImage* **109**, 418–428 (2015).

60. Takaso, H. et al. The effect of delayed auditory feedback on activity in the temporal lobe while speaking: a positron emission tomography study. *J. Speech Lang. Hear Res.* **53**, 226–236 (2010).

61. Vaquero, L. et al. The left, the better: white-matter brain integrity predicts foreign language imitation ability. *Cereb. Cortex* **4**, 2–12 (2016).

62. Kronfeld-Duenias, V. et al. Dorsal and ventral language pathways in persistent developmental stuttering. *Cortex* **81**, 79–92 (2016).

63. Neef, N. E. et al. Left posterior-dorsal area 44 couples with parietal areas to promote speech fluency, while right area 44 activity promotes the stopping of motor responses. *NeuroImage* **142**, 628–644 (2016).

64. Chevillet, M. A. et al. Automatic phoneme category selectivity in the dorsal auditory stream. *J. Neurosci.* **33**, 5208–5215 (2013).

65. Markiewicz, C. J. & Bohland, J. W. Mapping the cortical representation of speech sounds in a syllable repetition task. *NeuroImage* **141**, 174–190 (2016).

66. Alho, J. et al. Early-latency categorical speech sound representations in the left inferior frontal gyrus. *NeuroImage* **129**, 214–223 (2016).

67. Du, Y. et al. Noise differentially impacts phoneme representations in the auditory and speech motor systems. *Proc. Natl Acad. Sci. USA* **111**, 7126–7131 (2014).

68. Correia, J. M., Jansma, B. M. B. & Bonte, M. Decoding articulatory features from fMRI responses in dorsal speech regions. *J. Neurosci.* **35**, 15015–15025 (2015).

69. Kanero, J. et al. How sound symbolism is processed in the brain: a study on Japanese mimetic words. *PLOS ONE* **9**, e97905 (2014).

70. Agnew, Z. K., McGettigan, C. & Scott, S. K. Discriminating between auditory and motor cortical responses to speech and nonspeech mouth sounds. *J. Cogn. Neurosci.* **23**, 4038–4047 (2011).

71. Krishnan, S. et al. Beatboxers and guitarists engage sensorimotor regions selectively when listening to the instruments they can play. *Cereb. Cortex* **28**, 4063–4079 (2018).

72. Lewis, J. W. et al. Cortical networks representing object categories and high-level attributes of familiar real-world action sounds. *J. Cogn. Neurosci.* **23**, 2079–2101 (2011).

73. Engel, L. R. et al. Different categories of living and non-living sound-sources activate distinct cortical networks. *NeuroImage* **47**, 1778–1791 (2009).

74. Lewis, J. W. et al. Distinct cortical pathways for processing tool versus animal sounds. *J. Neurosci.* **25**, 5148–5158 (2005).

75. Repp, B. H. & Su, Y.-H. Sensorimotor synchronization: a review of recent research (2006–2012). *Psychon. Bull. Rev.* **20**, 403–452 (2013).

76. Pfordresher, P. Q. et al. Brain responses to altered auditory feedback during musical keyboard production — an fMRI study. *Brain Res.* **1556**, 28–37 (2014).

77. Gaver, W. W. What in the world do we hear? An ecological approach to auditory event perception. *Ecol. Psychol.* **5**, 1–29 (1993).

78. Warren, W. H. & Verbrugge, R. R. Auditory perception of breaking and bouncing events: a case study in ecological acoustics. *J. Exp. Psychol. Hum. Percept. Perform.* **10**, 704–712 (1984).

79. Ortiz-Rios, M. et al. Widespread and opponent fMRI signals represent sound location in macaque auditory cortex. *Neuron* **93**, 971–983 (2017).

80. Poirier, C. et al. Auditory motion-specific mechanisms in the primate brain. *PLOS Biol.* **15**, e2001379 (2017).

81. Fiehler, K. et al. Neural correlates of human echolocation of path direction during walking. *Multisens. Res.* **28**, 195–226 (2015).

82. Callan, A., Callan, D. E. & Ando, H. Neural correlates of sound externalization. *NeuroImage* **66**, 22–27 (2013).

83. Ceravolo, L., Frühholz, S. & Grandjean, D. Proximal vocal threat recruits the right voice-sensitive auditory cortex. *Soc. Cogn. Affect. Neurosci.* **11**, 793–802 (2016).

84. Ahveninen, J. et al. Evidence for distinct human auditory cortex regions for sound location versus identity processing. *Nat. Commun.* **4**, 615–619 (2013).

85. Zündorf, I. C., Lewald, J. & Karnath, H.-O. Testing the dual-pathway model for auditory processing in human cortex. *NeuroImage* **124**, 672–681 (2016).

86. Brungart, D. S. & Simpson, B. D. Within-ear and across-ear interference in a cocktail-party listening task. *J. Acoust. Soc. Amer.* **112**, 2985–2995 (2002).

87. Phillips, D. P. et al. Acoustic hemifields in the spatial release from masking of speech by noise. *J. Am. Acad. Audiol.* **14**, 518–524 (2003).

88. Mummery, C. J. et al. Functional neuroimaging of speech perception in six normal and two aphasic subjects. *J. Acoust. Soc. Amer.* **106**, 449–457 (1999).

89. Cohen, L. et al. Distinct unimodal and multimodal regions for word processing in the left temporal cortex. *NeuroImage* **23**, 1256–1270 (2004).

90. Scott, S. K., McGettigan, C. & Eisner, F. A little more conversation, a little less action — candidate roles for the motor cortex in speech perception. *Nat. Rev. Neurosci.* **10**, 295–302 (2009).

91. Zatorre, R. J., Belin, P. & Penhune, V. B. Structure and function of auditory cortex: music and speech. *Trends Cogn. Sci.* **6**, 37–46 (2002).

92. Zatorre, R. J. & Belin, P. Spectral and temporal processing in human auditory cortex. *Cereb. Cortex* **11**, 946–953 (2001).

93. Poeppel, D. The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. *Speech Commun.* **41**, 245–255 (2003).

94. Scott, S. K. et al. Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* **123**, 2400–2406 (2000).

95. Wise, R. J. S. et al. Separate neural subsystems within 'Wernicke's area'. *Brain* **124**, 83–95 (2001).

96. Winer, J. A. et al. Auditory thalamocortical transformation: structure and function. *Trends Neurosci.* **28**, 255–263 (2005).

97. Bizley, J. K. in *Conn's Translational Neuroscience* (ed. Conn, M. P.) 579–598 (Elsevier, 2017).

98. Chechik, G. et al. Reduction of information redundancy in the ascending auditory pathway. *Neuron* **51**, 359–368 (2006).

99. Goldstein, J. L. Auditory nonlinearity. *J. Acoust. Soc. Amer.* **41**, 676–699 (1967).

100. Fuchs, P. A., Glowatzki, E. & Moser, T. The afferent synapse of cochlear hair cells. *Curr. Opin. Neurobiol.* **13**, 452–458 (2003).

101. Harms, M. P. & Melcher, J. R. Sound repetition rate in the human auditory pathway: representations in the waveshape and amplitude of fMRI activation. *J. Neurophysiol.* **88**, 1433–1450 (2002).

102. Purcell, D. W. et al. Human temporal auditory acuity as assessed by envelope following responses. *J. Acoust. Soc. Amer.* **116**, 3581–3593 (2004).

103. Taylor, W. R. & Smith, R. G. The role of starburst amacrine cells in visual signal processing. *Vis. Neurosci.* **29**, 73–81 (2012).

104. Leff, A. P. et al. Impaired reading in patients with right hemianopia. *Ann. Neurol.* **47**, 171–178 (2000).

105. Coslett, H. B., Brashear, H. R. & Heilman, K. M. Pure word deafness after bilateral primary auditory cortex infarcts. *Neurology* **34**, 347–352 (1984).

106. Ulanovsky, N., Las, L. & Nelken, I. Processing of low-probability sounds by cortical neurons. *Nat. Neurosci.* **6**, 391–398 (2003).

107. Polterovich, A., Jankowski, M. M. & Nelken, I. Deviance sensitivity in the auditory cortex of freely moving rats. *PLOS ONE* **13**, e0197678 (2018).

108. Yao, J. D., Bremen, P. & Middlebrooks, J. C. Emergence of spatial stream segregation in the ascending auditory pathway. *J. Neurosci.* **35**, 16199–16212 (2015).

109. Slutsky, D. A. & Recanzone, G. H. Temporal and spatial dependency of the ventriloquism effect. *Neuroreport* **12**, 7–10 (2001).

110. Chen, Y., Repp, B. H. & Patel, A. D. Spectral decomposition of variability in synchronization and continuation tapping: comparisons between auditory and visual pacing and feedback conditions. *Hum. Mov. Sci.* **21**, 515–532 (2002).

111. Kaas, J. H. & Hackett, T. A. Subdivisions of auditory cortex and levels of processing in primates. *Audiol. Neurotol.* **3**, 73–85 (1998).