



## BIROn - Birkbeck Institutional Research Online

Jasmin, Kyle and Lima, C.F. and Scott, S.K. (2019) Understanding rostral-caudal auditory cortex contributions to auditory perception. *Nature Reviews Neuroscience* 20 , pp. 425-434. ISSN 1471-003X.

Downloaded from: <https://eprints.bbk.ac.uk/id/eprint/26151/>

*Usage Guidelines:*

Please refer to usage guidelines at <https://eprints.bbk.ac.uk/policies.html> or alternatively contact [lib-eprints@bbk.ac.uk](mailto:lib-eprints@bbk.ac.uk).

1  
2  
3 **Opinion**

4 **Understanding rostral-caudal auditory cortex contributions to auditory perception.**

5 Kyle Jasmin<sup>1,2\*</sup>, César F. Lima<sup>1,3</sup>, and Sophie K. Scott<sup>1\*</sup>

6 1. Institute of Cognitive Neuroscience, University College London, London, UK.

7 2. Department of Psychological Sciences, Birkbeck University of London, London, UK.

8 3. Instituto Universitário de Lisboa (ISCTE-IUL), Lisboa, Portugal.

9  
10 \*email: [sophie.scott@ucl.ac.uk](mailto:sophie.scott@ucl.ac.uk); [k.jasmin@bbk.ac.uk](mailto:k.jasmin@bbk.ac.uk)

11  
12 Abstract| There are functional and anatomical distinctions between the neural systems involved  
13 in the recognition of sounds in the environment and those involved in the sensorimotor guidance  
14 of sound production and the spatial processing of sound. Evidence for the separation of these  
15 processes has historically come from disparate literatures on the perception and production of  
16 speech, music and other sounds. More recent evidence indicates that there are computational  
17 distinctions between rostral and caudal primate auditory cortex that may underlie functional  
18 differences in auditory processing. These functional differences may originate from differences  
19 in the response times and temporal profiles of neurons in the rostral and caudal auditory cortex,  
20 suggesting that computational accounts of primate auditory pathways should focus on the  
21 implications of these temporal response differences.

22 **[H1] Introduction**

23 The primate visual and auditory perceptual systems are equally complex, but approaches to  
24 untangling their complexity have differed: whereas models of visual processing have been often  
25 examined in a domain-general way (by measuring neural responses to basic visual features, for  
26 example), models of the auditory system have tended to focus on specific domains of auditory  
27 processing, such as the perception of intelligible speech and language<sup>1-3</sup>, the perception of  
28 linguistic and emotional prosody<sup>4,5</sup>, and the perception and production of music<sup>6,7</sup>. Studying  
29 these specific domains has proved useful for determining the functional properties of auditory  
30 cortex and it is arguable that beginning with such approaches was in some ways necessary. For

31 instance, the functional organization of macaque auditory cortex into a rostral ‘recognition’  
32 pathway and a caudal ‘spatial’ pathway was not apparent when simple tones (designed to be  
33 analogous to simple visual features) were used as stimuli<sup>8</sup>. It was only when the vocal calls of  
34 monkey conspecifics were used that these properties became obvious<sup>9</sup>. Furthermore, there is also  
35 strong evidence that different kinds of auditory information are represented in distinct parts of  
36 the brain: for example, stroke can rob someone of the ability to understand music while  
37 preserving functions such as the comprehension of speech and other sounds<sup>10</sup>. Nevertheless,  
38 domain-specific approaches to understanding audition cannot (or do not aim to) account for the  
39 perception and processing of sounds outside these domains (such as impact sounds, which are  
40 neither vocal nor musical). What is therefore needed is a domain-general model in which there  
41 are multiple interacting computations, such as those that have been proposed for vision<sup>e.g. 11</sup>.

42         Recent developments in auditory neuroscience have begun to reveal candidate  
43 organisational principles for the processing of sound in the primate brain<sup>12-14</sup>. In this article, we  
44 argue that these organisational principles can be used to develop more computationally driven,  
45 domain-general models of cortical auditory processing. Previous reviews on auditory processing  
46 have characterized the involvement of rostral and caudal pathways with specific auditory and  
47 linguistic domains<sup>1-7</sup>. Other accounts have posited the relationship of these pathways to attention  
48<sup>15,16</sup> or described their role in perceiving auditory objects<sup>17</sup>. Our purpose here is rather different.  
49 We will describe and synthesise recent findings of auditory neuroscience studies that have used  
50 neuroanatomical analyses, electrocorticography (ECoG) and functional MRI (fMRI) in humans  
51 and monkeys, with the aim of setting out a domain-general functional account of the primate  
52 auditory cortex. The model that we propose is based on rostro–caudal patterns of intracortical  
53 and extracortical connectivity in the auditory cortex, the differential temporal response properties  
54 of rostral and caudal cortical fields and task-related functional engagement of rostral and caudal  
55 regions of the auditory cortex.

56

## 57 **[H1] Auditory anatomical organisation**

58 In audition, the signal carried by the auditory nerve is deconstructed into different kinds of  
59 informational features, which are represented in parallel in the ascending auditory pathway (Box  
60 1). Within these representations, some general organisational principles are apparent. Tonotopy —  
61 in which the frequency information in sound is represented across a spatial array — is first

62 established in the cochlea and is preserved along the entire ascending auditory pathway<sup>18</sup>. In  
63 addition, other acoustic features — such as sound onsets and offsets, temporal regularities relating  
64 to pitch, and spatial location — are computed from the cochlear nucleus onwards<sup>18</sup>. Thus, there is  
65 intense complexity in the sub-cortical processing of sound and this complexity (Box 2) is preserved  
66 even as the temporal detail of the sound representations decreases (Box 1). Following this  
67 subcortical processing, the medial geniculate body (auditory thalamus) projects to the cortex  
68 (which also makes strong connections back to subcortical nuclei; Fig. 1a).

69         The primate auditory cortex is organised, anatomically, in a rostral–caudal orientation, with  
70 three core primary fields surrounded by belt and parabelt fields, in a roughly concentric form. The  
71 tonotopic organisation seen in the ascending auditory pathway is seen within the core fields, with  
72 three different tonotopic gradients seen across the three core fields<sup>8,19</sup>. Connectivity within the  
73 ‘core’ auditory cortex also maintains a rostral–caudal axis, with greater connectivity between  
74 adjacent core auditory regions than between non-adjacent core fields (Fig. 1b)<sup>12</sup>. This rostral–  
75 caudal organisation is also seen in the connections between the auditory thalamus and the rostral  
76 and caudal core auditory fields: A1 and R both receive the vast majority of their inputs from the  
77 ventral medial geniculate body, whereas the rostral-most core field, RT, receives a greater  
78 proportion of inputs from the postero-dorsal medial geniculate body (Figure 1a)<sup>13</sup>. Rostral–caudal  
79 differences extend into the thalamo–cortical connectivity of rostral belt and parabelt areas. The  
80 rostral belt area RTp, lying directly rostral to RT, receives most of its inputs from the postero-  
81 dorsal, ventral and medial fields within the medial geniculate body, whereas the rostral superior  
82 temporal gyrus, lying lateral to RTp in the parabelt is more strongly connected to the medial  
83 pulvinar and supragenulate (Sg)–limitans (Lim) complex<sup>13</sup>.

84         This rostral–caudal organization of anatomy and connectivity has been taken as  
85 contributing evidence to support the idea that the nature of processing in rostral and caudal auditory  
86 cortex may be qualitatively distinct. For instance, it has been suggested that rostral projections and  
87 fields may be more likely (than caudal projections) to play a fundamental role in the integration of  
88 audiovisual information, since they are more strongly anatomically connected to thalamic nuclei  
89 that process visual information as well as sound (the medial pulvinar and Sg and Lim thalamic  
90 nuclei; Box 1)<sup>13</sup>. Caudal areas, on the other hand, are proposed to be involved with processing  
91 audio-somatosensory stimuli, responding both to sounds (such as clicks) and to facial  
92 somatosensory stimulation<sup>20,21</sup> and may mediate the roles of facial somato-sensation and sound

93 processing in the control of articulation<sup>22</sup>. In support of this proposal, caudal belt regions do not  
94 receive inputs from visual thalamus, but do show (in addition to auditory thalamus connectivity)  
95 input from the somatosensory thalamus<sup>21</sup>.

96

## 97 **[H1] Auditory response properties**

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

108 Nevertheless, important rostral–caudal differences can be seen in the speeds of neural  
109 responses and neural sensitivities to the structure of sounds over time<sup>24</sup>. There are rostral–caudal  
110 differences in response latency in both core and belt areas<sup>25</sup>: caudal core area A1 shows a faster  
111 median response to sounds (within 20ms of onset) than the more rostral area R (33ms)<sup>14</sup> (Figure  
112 2) and Caudo-medial belt areas have been shown to have an average response latency of around  
113 10ms, even faster than core areas<sup>26</sup>. A1 also tracks fast acoustic amplitude modulations (with a  
114 duty cycle on the order of 20-30ms) more accurately than the more rostral core area R, which can  
115 only track slower amplitude modulations (with a duty cycle on the order of 100ms and above)<sup>14</sup>.  
116 Neurons in area R saturate in their response at lower frequencies than those in A1 (Fig. 2),  
117 indicating that neurons in area R lose synchrony at lower rates than those in A1 which can continue  
118 to synchronize to faster rates of amplitude modulation<sup>14</sup>. In other words, the caudal core auditory  
119 cortex responds quickly to sound onsets, and tracks fast amplitude modulations accurately,  
120 whereas rostral auditory cortex responds more slowly to the starts of sounds and more accurately  
121 tracks slower amplitude modulations (such as those found in speech).

122 Although these rostral–caudal temporal processing differences are not completely distinct,  
123 and the similarities in response properties of core areas should not be ignored<sup>27</sup>, we hypothesise

124 that they may relate to important functional differences at higher levels of the auditory cortex. The  
125 faster and more precise temporal response in caudal A1 suggests that caudal auditory fields may  
126 more accurately compute certain aspects of sound sequences than more rostral fields. For example,  
127 the perception of rhythm is based on perceived beat onsets in sounds. The finer temporal acuity of  
128 caudal areas may make them better at tracking and coding these perceived onsets than the rostral  
129 auditory cortex, which also has a poorer resolution of different amplitude modulation frequencies  
130 (perceived beat onsets have been linked to amplitude onsets, as opposed to amplitude offsets<sup>28</sup>).  
131 We hypothesise that this difference in temporal acuity may also make caudal auditory cortex  
132 suitable for performing computations that guide actions, which need to occur quickly if they are  
133 to be of any utility in the control of movement. There is evidence that engaging the motor system  
134 in an auditory perception task does indeed increase the temporal acuity of responses, which may  
135 reflect the enhanced involvement of rostral auditory fields. For instance, it is known that  
136 participants are more accurate at tracking changes in auditory intervals when they are tapping  
137 along than when they are passively listening, and we would suggest that this reflects differential  
138 recruitment of caudal auditory sensory motor systems which are recruited by coordinating actions  
139 with sounds<sup>29</sup>.

140 By contrast to the brisk onset responses seen in caudal fields, we suggest that the slow  
141 onset times seen in rostral fields may reflect processes that are slower and that entail feedforward  
142 and feedback patterns of connectivity. Circuits mediating hierarchical perceptual processing and  
143 recognition processes, for example, tend to be slower in their responses, which reflects the time  
144 courses of prediction and integration of incoming perceptual information with prior experience  
145 (for example, the use of context in understanding)<sup>30</sup>. Indeed, auditory recognition processes can  
146 be relatively slow: in humans, electrophysiological studies show that the earliest neural correlates  
147 of auditory semantic processing can be seen about 200ms after stimulus presentation, and continue  
148 to unfold over a further several hundred milliseconds<sup>30</sup>. We therefore suggest that rostral auditory  
149 cortex areas may be well suited to a role in such processes.

150 Further evidence for the distinct temporal response properties of rostral and caudal auditory  
151 cortex comes from human studies. In a recent ECoG experiment<sup>31</sup>, cortical recordings were  
152 obtained as 27 participants heard spoken sentences. An unsupervised learning approach clustered  
153 the neural responses according to their temporal profiles. The results revealed large rostral–caudal  
154 differences: caudal fields responded quickly and strongly to the onsets of sentences (with

155 additional onset responses occurring after gaps in speech longer than 200ms). Rostral fields, by  
156 contrast, showed much weaker onset responses and produced slower and more sustained  
157 responses. This difference (Fig. 2) supports the idea that there are computational differences  
158 between rostral and caudal auditory cortex fields in terms of basic acoustic processing, with caudal  
159 fields being more sensitive to onsets (and hence the temporal characteristics of sounds) and rostral  
160 fields being more sensitive to the spectrotemporal information conveyed over the whole sequence  
161 of a sound. Indeed, pure tones — which do not contain changing structure over time — produced  
162 only caudal onset responses. The results were seen over all stimuli (including nearly 500 natural  
163 sentences, and also single syllables), and did not depend on the linguistic properties of the sentence  
164 or the phonetic properties of the speech sounds, indicating that this may represent a more global  
165 rostral/caudal distinction in temporal response characteristics. We note, however, that ECoG study  
166 participants are almost always patients with intractable epilepsy who are on medication and may  
167 have suffered brain tissue damage or trauma. Thus the results from such studies should be  
168 interpreted cautiously and corroborated with evidence obtained with other techniques.

169 Human functional imaging has revealed processing differences in rostral and caudal  
170 auditory areas in humans that are in line with the monkey electrophysiology and human ECoG  
171 findings discussed above. In one study, participants were presented with a variety of different kinds  
172 of sounds (including speech, emotional vocalizations, animal cries, musical instruments and tool  
173 sounds). Their cortical responses (measured with fMRI) were analysed with respect to the spectro-  
174 temporal features of the sounds<sup>32</sup>. The presentation of sounds in which there were fast modulations  
175 of the amplitude envelope (that is, the changes in amplitude of a sound over time) but slower  
176 spectral modulations (that is, changes in the large-scale distribution of the frequency content of the  
177 sounds, such as those that characterise formants in speech) led to an enhanced response in medial  
178 regions caudal to Heschl's gyrus, (the major anatomical landmark for primary auditory fields in  
179 the human brain), whereas sounds that contained more detailed spectral information and broader  
180 amplitude envelope modulations were associated with responses in regions rostral to Heschl's  
181 gyrus and in the rostral superior temporal gyrus (STG). This pattern was replicated in a second  
182 fMRI study that examined responses to environmental sounds, speech and music<sup>33</sup>: caudal auditory  
183 fields responded preferentially to fast amplitude envelope modulations and slower spectral  
184 modulations, whereas rostral fields responded preferentially to faster spectral modulations and  
185 slow amplitude envelope modulations. As in the experiments in monkeys, these response profiles

186 suggest that caudal and rostral regions may be involved in distinct computations: rostral fields may  
187 process information conveyed in the spectral detail of a sound, whereas caudal fields may process  
188 information conveyed via the amplitude envelope. Below, we will examine the more specific  
189 functional properties of rostral and caudal auditory regions that these neuronal differences may  
190 indicate.

## 191 **[H1] Rostral auditory processing**

### 192 *[H2] Recognition processes*

193 Human speech is a perfect example of a spectrally complex sound. Comprehending speech  
194 requires the auditory system to grapple with dynamic changes in the spectral profile and, from this,  
195 recognize meaningful units of sound (such as phonemes, words, and grammatical and prosodic  
196 structures). Given the rostral auditory cortex's proposed role in processing spectrally complex  
197 sounds, it is unsurprising that activity levels in rostral auditory fields are highly sensitive to the  
198 intelligibility of speech<sup>34</sup> and that sub-fields within rostral STG respond to specific components of  
199 intelligible speech such as syntactic structures<sup>35,36</sup>.

200 Based on a review of the literature, it has been argued that there is a caudal–rostral  
201 hierarchical processing gradient for speech (mainly in the left cortical hemisphere), in which the  
202 most basic acoustic processing takes place in the primary auditory cortex and the complexity of  
203 processing increases as the information progresses rostrally, with the processing of high-order  
204 lexical and semantic structure taking place near the temporal pole<sup>37</sup>. This proposed hierarchy  
205 mirrors the gradient in cortical thickness measurements in temporal areas: the cortex is thinner and  
206 there are fewer feedback connections crossing cortical layers near the primary auditory cortex,  
207 whereas the cortex is thicker and has a higher ratio of feedback connections (those from deeper  
208 cortical layers to superficial cortical layers) to feedforward connections near the temporal pole<sup>38</sup>.  
209 The greater number of connections across layers has typically been assumed to be linked to greater  
210 processing complexity<sup>38</sup>. Furthermore, physiological studies in non-human primates have shown  
211 that rostral STG auditory areas exhibit more inhibitory responses than excitatory responses and  
212 that the latencies of these responses are longer than they are in more caudal areas: properties  
213 indicative of a higher position in a hierarchical processing stream<sup>39</sup>. It is also the case that rostral  
214 superior temporal lobe responses to speech appear to be malleable and sensitive to the effects of

215 prediction and context: whereas mid STG fields are unaffected by sentence expectations, rostral  
216 auditory areas respond selectively based on the expected or violated sentence endings<sup>40</sup>. Notably,  
217 this more context sensitive response is mediated by input from the larger language network, and  
218 is associated with specific connectivity between rostral auditory areas and ventral frontal cortex  
219 fields<sup>40,41</sup>.

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

## 229 ***[H2] Parallel processing of multiple auditory objects***

230 In normal environments, we frequently hear multiple auditory objects simultaneously (at the time  
231 of writing, for example, we can hear a car alarm and footsteps in the corridor, in addition to the  
232 sounds of our own typing). We know that unattended auditory information can disrupt  
233 performance in behavioural tasks requiring speech production or holding verbal information in  
234 working memory, which suggests that unattended auditory objects are being processed (to some  
235 extent) for meaning, in parallel with attended auditory information<sup>44</sup>. The ascending auditory  
236 pathways are essential for forming representations of auditory objects and their associated spatial  
237 locations (Box 1) and rostral cortical auditory fields appear to be capable of representing multiple  
238 parallel auditory objects, only one of which forms the currently attended signal<sup>45,46</sup>. Studies of  
239 ‘masked’ speech, in which a target speech signal is heard against a simultaneous competing sound,  
240 indicate that when a competing sound is more speech-like, it elicits a greater neural response in  
241 rostral auditory fields<sup>47,48</sup>. This response occurs in addition to the activation associated with the  
242 content of the attended speech<sup>47</sup> (which may include self-produced speech<sup>48</sup>), suggesting that the  
243 computational processes taking place in the rostral auditory cortex must be flexible enough to  
244 process (and recognise aspects of) multiple unattended auditory objects. This flexibility must

245 permit the processing of multiple parallel sources of auditory information for a wide variety of  
246 possible kinds of sound, as well as the switching of attentional focus between them. Such switching  
247 may occur on the basis of intention and/or when information in an unattended stream starts to  
248 compete for resources with the attended stream<sup>49</sup>. Such parallel processing must therefore be fast,  
249 plastic and highly state-dependent.

## 250 **[H1] Caudal auditory processing**

### 251 *[H2] Sensorimotor and spatial computations*

252 As discussed above, caudal auditory fields show precise and rapid responses to sound onsets and  
253 fluctuations. We suggest that this makes them ideal for guiding motor responses to sounds in the  
254 environment or to self-produced sounds, especially those that require rapid action. Speech  
255 production is, of course, a motor action that requires tight temporal and spatial control<sup>50</sup>. Caudal  
256 auditory cortical fields have been shown many times to be recruited during speech production<sup>51-</sup>  
257 <sup>55</sup>, whereas the activity of rostral auditory fields is suppressed during articulation (relative to its  
258 activity when hearing speech)<sup>56,57</sup>. This motor-related caudal auditory activity is enhanced when a  
259 talker, for example, alters their voice to match specific pitches<sup>58</sup>, compensates for an  
260 experimentally induced altered shift in their perceived vocal pitch<sup>59</sup> or speaks (usually with  
261 significant disfluencies) while being presented with delayed auditory feedback<sup>60</sup>. Superior  
262 auditory-motor abilities have been shown to correlate with neural measures in pathways connected  
263 to caudal auditory fields. For example, the arcuate fasciculus, a white matter tract that projects  
264 from caudal temporal and parietal cortex to the frontal lobe, shows greater leftward lateralization  
265 in terms of volume and increased integrity (measured with fractional anisotropy) in people who  
266 are better at imitating foreign accents<sup>61</sup>. Conversely, difficulties with speech production (such as  
267 stammering) have been linked to abnormalities in pathways connected to caudal auditory  
268 fields<sup>62,63</sup>.

269 Rostral auditory streams support recognition processes under normal listening conditions  
270 (see above); however, caudal areas do seem to play a limited role in recognition processes. Caudal  
271 areas are recruited only during some specific kinds of perceptual task, including those requiring  
272 sublexical units (such as phonemes)<sup>64-67</sup> and phonetic features<sup>68</sup> of speech to be accessed, motor-  
273 related semantic features to be processed (as is the case for Japanese onomatopoeic words<sup>69</sup>), or

274 the passive perception of non-speech mouth sounds that ‘sound do-able’ (can be matched to an  
275 action) <sup>22,70</sup>. It is also important to note that when auditory recognition processes require an  
276 emphasis on the way that a sound is made — for example, when beat boxers hear unfamiliar  
277 examples of expert beat boxing<sup>71</sup> or people listen to sounds produced by human actions <sup>72,73</sup> (such  
278 as the sounds made by hand tools<sup>74</sup>) — caudal auditory areas are recruited and form part of a wider  
279 sensorimotor network.

280 Although speech is usually considered the prototypical sound-related action, the  
281 audiomotor integration of other types of sounds also relies on caudal fields. Musical performance,  
282 for example, requires precise cortical responses to sound in order to guide accurate motor  
283 production<sup>75</sup>. Although there is much less published research on the neuroscience of music than  
284 there is on speech, effects similar to those observed for speech (such as an enhanced caudal  
285 auditory cortex response) are seen when study participants attempt to perform music while  
286 receiving perceptual feedback that is altered<sup>76</sup>. Action-related sounds also guide many other  
287 movements: for example, the rising resonance frequency as glass of water is filled indicates when  
288 to stop pouring. Similarly, the sound an egg makes when it cracks, and many other actions-related  
289 sounds, require precise responses <sup>77,78</sup> and startle reflexes to loud sounds entail an immediate  
290 orientation to the perceived sound location. Although we know of no published studies on motor  
291 guidance in response to such sounds, we predict that they should also recruit caudal auditory cortex  
292 fields.

293 Movement is, of course, closely linked with space. The bias for responses to the onsets of  
294 sounds in caudal auditory cortex<sup>31</sup>, combined with the capacity of caudal areas to produce fast and  
295 fine temporal response to sounds, could make these areas suitable for locating sounds in space<sup>79</sup>  
296 and guiding and processing movement<sup>80</sup> and navigation (as described in the following studies)  
297 accordingly. Recent evidence from fMRI studies that have used binaural and 3D sound  
298 presentation paradigms supports this: blind human participants showed increased caudal temporal  
299 (and parietal) responses to echoes when listening to recordings that were binaural (and therefore  
300 contained the information necessary for echolocation) than they did when the recordings were  
301 monaural<sup>81</sup>. Similarly, in sighted people, sounds presented binaurally to create the illusion of a  
302 source existing outside the head activate caudal pathways more strongly than those presented  
303 monaurally (which lack information necessary to calculate a spatial origin and therefore appear  
304 to originate inside the head<sup>82</sup>). Caudal superior temporal activity is also modulated by varying the

305 perceived location of a sound in space, as indicated by its direction<sup>82</sup> and proximity to the head<sup>83</sup>.  
306 However, single cell recording studies in caudal belt fields find partially segregated responses to  
307 temporal features and spatial location, which suggests that two independent streams may  
308 contribute to these sensorimotor and spatial processes<sup>26</sup>.

309 FMRI is an inherently correlational technique, but these computational distinctions are also  
310 supported by causal evidence obtained from a transcranial magnetic stimulation (TMS) study and  
311 a patient study: whereas transient TMS applied to the rostral auditory cortex delayed reaction times  
312 for judgements concerning sound identity more than it affected those related to sound location,  
313 similarly disrupting the caudal auditory cortex delayed judgments of sound location more than it  
314 did sound identity<sup>84</sup>. Similarly, stroke damage to rostral areas affects sound identification, whereas  
315 damage to caudal auditory cortex impairs location judgments<sup>85</sup>.

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

325

## 326 **[H1] Future directions**

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

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

344 We also still do not know exactly how multiple auditory objects are represented, processed  
345 and selected between in rostral auditory fields, or precisely what kinds of auditory information are  
346 used to guide action. Is it really the case that the fine temporal sensitivity of caudal fields is  
347 matched by a weaker reliance on spectral cues<sup>90</sup> or is the system more complex than this? When  
348 we understand what aspects of sounds are represented at distinct levels of both cortical and  
349 subcortical processing, the corresponding acoustic profiles and the resulting functional responses  
350 (that is, how they are used), we will have moved closer to a computational model of primate  
351 hearing.

352  
353 Several previous papers have put forward models of the properties of distinct auditory processing  
354 streams<sup>1-3,5,7,90,91</sup>. Our proposal is distinct in that we are trying to synthesize across a wider range  
355 of auditory domains than the previous domain-specific models, and we have taken temporal  
356 response properties of neurons to sound as a feature to distinguish the two candidate systems'  
357 computational differences. A couple of previous approaches have used temporal processing as a  
358 way of distinguishing differences in auditory processing; however, both focussed on the temporal  
359 characteristics of sounds and used this as a way of hypothesizing candidate processing differences  
360 between the left and right auditory fields. One model<sup>92</sup> suggested that, by analogy with the  
361 construction of spectrograms, the left auditory fields had good temporal resolution and poor  
362 spectral resolution, whereas the analogous regions on the right had poor temporal resolution.  
363 Another<sup>93</sup> specifically suggested that the left auditory fields sampled sounds at a faster rate than  
364 the right auditory fields, with a general model of 'window size' being shorter in the left and longer  
365 on the right. Both of these approaches aim to account for hemispheric asymmetries in speech and  
366 sound processing by positing selective processing of particular acoustic characteristics. However,

367 we believe that the evidence for such specificity of acoustic processing is sparse<sup>50</sup>. Other previous  
368 reviews and meta-analyses have focused only on the functions of auditory pathways: that is, how  
369 they interact with factors such as attention<sup>15,16</sup> or their roles in segmenting continuous sound into  
370 discrete auditory objects<sup>17</sup>. What we are suggesting in this article, which diverges from previous  
371 accounts, is that the temporal response characteristics of the rostral and caudal auditory cortex  
372 fields are distinct and may underlie computational differences that give rise to previously observed  
373 functional differences.

#### 374 **[H1] Concluding remarks**

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

396

#### 397 **References**

- 398 1. Rauschecker, J. P. & Scott, S. K. Maps and streams in the auditory cortex: nonhuman  
399 primates illuminate human speech processing. *Nat Neurosci* **12**, 718–724 (2009).
- 400 2. Scott, S. K. & Johnsrude, I. S. The neuroanatomical and functional organization of  
401 speech perception. *Trends in Neurosciences* **26**, 100–107 (2003).
- 402 3. Hickok, G. & Poeppel, D. Dorsal and ventral streams: a framework for understanding  
403 aspects of the functional anatomy of language. *Cognition* **92**, 67–99 (2004).
- 404 4. Sammler, D., Grosbras, M.-H., Anwander, A., Bestelmeyer, P. E. G. & Belin, P. Dorsal  
405 and Ventral Pathways for Prosody. *Current Biology* **25**, 3079–3085 (2015).
- 406 5. Schirmer, A. & Kotz, S. A. Beyond the right hemisphere: brain mechanisms mediating  
407 vocal emotional processing. *Trends in Cognitive Sciences* **10**, 24–30 (2006).
- 408 6. Zatorre, R. J., Chen, J. L. & Penhune, V. B. When the brain plays music: auditory–motor  
409 interactions in music perception and production. *Nat Rev Neurosci* **8**, 547–558 (2007).
- 410 7. Alain, C., Arnott, S. R., Hevenor, S., Graham, S. & Grady, C. L. "What" and 'where' in  
411 the human auditory system. *Proc Natl Acad Sci USA* **98**, 12301–12306 (2001).
- 412 8. Rauschecker, J. P. Processing of Complex Sounds in the Auditory Cortex of Cat,  
413 Monkey, and Man. *Acta Oto-Laryngologica* **117**, 34–38 (1997).
- 414 9. Rauschecker, J. P. & Tian, B. Mechanisms and streams for processing of "what" and  
415 "where" in auditory cortex. *Proc Natl Acad Sci USA* **97**, 11800–11806 (2000).
- 416 10. Rosemann, S., Brunner, F., Kastrop, A. & Fahle, M. Musical, visual and cognitive  
417 deficits after middle cerebral artery infarction. *eNeurologicalSci* **6**, 25–32 (2017).
- 418 11. Kravitz, D. J., Saleem, K. S., Baker, C. I. & Mishkin, M. A new neural framework for  
419 visuospatial processing. *Nat Rev Neurosci* **12**, 1–14 (2011).
- 420 12. Scott, B. H. *et al.* Intrinsic Connections of the Core Auditory Cortical Regions and  
421 Rostral Supratemporal Plane in the Macaque Monkey. *Cerebral Cortex* **7**, 809–840  
422 (2015).
- 423 13. Scott, B. H. *et al.* Thalamic connections of the core auditory cortex and rostral  
424 supratemporal plane in the macaque monkey. *J. Comp. Neurol.* **525**, 3488–3513 (2017).
- 425 14. Scott, B. H., Malone, B. J. & Semple, M. N. Transformation of Temporal Processing  
426 Across Auditory Cortex of Awake Macaques. *Journal of Neurophysiology* **105**, 712–730  
427 (2011).
- 428 15. Arnott, S. R. & Alain, C. The auditory dorsal pathway: Orienting vision. *Neuroscience*  
429 *and Biobehavioral Reviews* **35**, 2162–2173 (2011).
- 430 16. Alho, K., Rinne, T., Herron, T. J. & Woods, D. L. Stimulus-dependent activations and  
431 attention-related modulations in the auditory cortex: A meta-analysis of fMRI studies.  
432 *Hearing Research* **307**, 29–41 (2014).
- 433 17. Bizley, J. K. & Cohen, Y. E. The what, where and how of auditory-object perception.  
434 *Nat Rev Neurosci* **14**, 693–707 (2013).
- 435 18. Young, E. D. & Oertel, D. in *The Synaptic Organization of the Brain* 125–164 (Oxford  
436 University Press, 2004). doi:10.1093/acprof:oso/9780195159561.003.0004
- 437 19. Kaas, J. H. & Hackett, T. A. Subdivisions of auditory cortex and processing streams in  
438 primates. *Proc Natl Acad Sci USA* **97**, 11793–11799 (2000).
- 439 20. Smiley, J. F. *et al.* Multisensory convergence in auditory cortex, I. Cortical connections  
440 of the caudal superior temporal plane in macaque monkeys. *J. Comp. Neurol.* **502**, 894–  
441 923 (2007).
- 442 21. Hackett, T. A. *et al.* Multisensory convergence in auditory cortex, II. Thalamocortical  
443 connections of the caudal superior temporal plane. *J. Comp. Neurol.* **502**, 924–952

- 444 (2007).
- 445 22. Warren, J. E., Wise, R. J. S. & Warren, J. D. Sounds do-able: auditory-motor  
446 transformations and the posterior temporal plane. *Trends in Neurosciences* **28**, 636–643  
447 (2005).
- 448 23. Dick, F. *et al.* In Vivo Functional and Myeloarchitectonic Mapping of Human Primary  
449 Auditory Areas. *Journal of Neuroscience* **32**, 16095–16105 (2012).
- 450 24. Rauschecker, J. P. Where, When, and How: Are they all sensorimotor? Towards a  
451 unified view of the dorsal pathway in vision and audition. *CORTEX* **98**, 262–268 (2018).
- 452 25. Camalier, C. R., D'Angelo, W. R., Sterbing-D'Angelo, S. J., la Mothe, de, L. A. &  
453 Hackett, T. A. Neural latencies across auditory cortex of macaque support a dorsal  
454 stream supramodal timing advantage in primates. *Proc Natl Acad Sci USA* **109**, 18168–  
455 18173 (2012).
- 456 26. Kuśmierek, P. & Rauschecker, J. P. Selectivity for space and time in early areas of the  
457 auditory dorsal stream in the rhesus monkey. *Journal of Neurophysiology* **111**, 1671–  
458 1685 (2014).
- 459 27. Smith, E. H. Temporal processing in the auditory core: transformation or segregation?  
460 *Journal of Neurophysiology* **106**, 2791–2793 (2011).
- 461 28. Scott, S. K. The point of P-centres. *Psychological Research Psychologische Forschung*  
462 **61**, 4–11 (1998).
- 463 29. Repp, B. H. & Keller, P. E. Adaptation to tempo changes in sensorimotor  
464 synchronization: Effects of intention, attention, and awareness. *The Quarterly Journal of*  
465 *Experimental Psychology Section A* **57**, 499–521 (2004).
- 466 30. Holcomb, P. J. & Neville, H. J. Auditory and Visual Semantic Priming in Lexical  
467 Decision: A Comparison Using Event-related Brain Potentials. *Language and Cognitive*  
468 *Processes* **5**, 281–312 (1990).
- 469 31. Hamilton, L. S., Edwards, E. & Chang, E. F. A Spatial Map of Onset and Sustained  
470 Responses to Speech in the Human Superior Temporal Gyrus. *Current Biology* **28**,  
471 1860–1871.e4 (2018).
- 472 32. Santoro, R. *et al.* Encoding of Natural Sounds at Multiple Spectral and Temporal  
473 Resolutions in the Human Auditory Cortex. *PLoS Comput Biol* **10**, e1003412–14 (2014).
- 474 33. Norman-Haignere, S., Kanwisher, N. G. & McDermott, J. H. Distinct Cortical Pathways  
475 for Music and Speech Revealed by Hypothesis-Free Voxel Decomposition. *Neuron* **88**,  
476 1281–1296 (2015).
- 477 34. Evans, S. *et al.* The Pathways for Intelligible Speech: Multivariate and Univariate  
478 Perspectives. *Cerebral Cortex* **24**, 2350–2361 (2014).
- 479 35. Agnew, Z. K., van de Koot, H., McGettigan, C. & Scott, S. K. Do sentences with  
480 unaccusative verbs involve syntactic movement? Evidence from neuroimaging.  
481 *Language, Cognition and Neuroscience* **29**, 1035–1045 (2014).
- 482 36. de Heer, W. A., Huth, A. G., Griffiths, T. L., Gallant, J. L. & Theunissen, F. E. The  
483 Hierarchical Cortical Organization of Human Speech Processing. *Journal of*  
484 *Neuroscience* **37**, 6539–6557 (2017).
- 485 37. Specht, K. Mapping a lateralization gradient within the ventral stream for auditory  
486 speech perception. *Front. Hum. Neurosci.* **7**, (2013).
- 487 38. Wagstyl, K., Ronan, L., Goodyer, I. M. & Fletcher, P. C. Cortical thickness gradients in  
488 structural hierarchies. *NeuroImage* **111**, 241–250 (2015).
- 489 39. Kikuchi, Y., Horwitz, B. & Mishkin, M. Hierarchical auditory processing directed

- 490 rostrally along the monkey's supratemporal plane. *J. Neurosci.* **30**, 13021–13030 (2010).
- 491 40. Tuennerhoff, J. & Noppeney, U. When sentences live up to your expectations.
- 492 *NeuroImage* **124**, 641–653 (2016).
- 493 41. Lyu, B., Ge, J., Niu, Z., Tan, L. H. & Gao, J.-H. Predictive Brain Mechanisms in Sound-
- 494 to-Meaning Mapping during Speech Processing. *Journal of Neuroscience* **36**, 10813–
- 495 10822 (2016).
- 496 42. Leaver, A. M. & Rauschecker, J. P. Cortical Representation of Natural Complex Sounds:
- 497 Effects of Acoustic Features and Auditory Object Category. *Journal of Neuroscience* **30**,
- 498 7604–7612 (2010).
- 499 43. Price, C., Thierry, G. & Griffiths, T. Speech-specific auditory processing: where is it?
- 500 *Trends in Cognitive Sciences* **9**, 271–276 (2005).
- 501 44. Beaman, C. P. & Jones, D. M. Irrelevant Sound Disrupts Order Information in Free
- 502 Recall as in Serial Recall. *The Quarterly Journal of Experimental Psychology Section A*
- 503 **51**, 615–636 (1998).
- 504 45. Scott, S. K. Auditory processing — speech, space and auditory objects. *Current Opinion*
- 505 *in Neurobiology* **15**, 197–201 (2005).
- 506 46. Zatorre, R. J. Sensitivity to Auditory Object Features in Human Temporal Neocortex.
- 507 *Journal of Neuroscience* **24**, 3637–3642 (2004).
- 508 47. Evans, S., McGettigan, C., Agnew, Z. K., Rosen, S. & Scott, S. K. Getting the Cocktail
- 509 Party Started: Masking Effects in Speech Perception. *Journal of Cognitive Neuroscience*
- 510 **28**, 483–500 (2016).
- 511 48. Meekings, S. *et al.* Distinct neural systems recruited when speech production is
- 512 modulated by different masking sounds. *The Journal of the Acoustical Society of*
- 513 *America* **140**, 8–19 (2016).
- 514 49. Brungart, D. S., Simpson, B. D., Ericson, M. A. & Scott, K. R. Informational and
- 515 energetic masking effects in the perception of multiple simultaneous talkers. *The Journal*
- 516 *of the Acoustical Society of America* **110**, 2527–2538 (2001).
- 517 50. McGettigan, C. & Scott, S. K. Cortical asymmetries in speech perception: what's wrong,
- 518 what's right and what's left? *Trends in Cognitive Sciences* **16**, 269–276 (2012).
- 519 51. Hickok, G. A functional magnetic resonance imaging study of the role of left posterior
- 520 superior temporal gyrus in speech production: implications for the explanation of
- 521 conduction aphasia. *Neuroscience Letters* **287**, 156–160 (2000).
- 522 52. Flinker, A. *et al.* Single-Trial Speech Suppression of Auditory Cortex Activity in
- 523 Humans. *Journal of Neuroscience* **30**, 16643–16650 (2010).
- 524 53. Agnew, Z. K., McGettigan, C., Banks, B. & Scott, S. K. Articulatory movements
- 525 modulate auditory responses to speech. *NeuroImage* **73**, 191–199 (2013).
- 526 54. Jasmin, K. M. *et al.* Cohesion and Joint Speech: Right Hemisphere Contributions to
- 527 Synchronized Vocal Production. *Journal of Neuroscience* **36**, 4669–4680 (2016).
- 528 55. Jasmin, K. *et al.* Overt social interaction and resting state in autism: core and contextual
- 529 neural features. *bioRxiv* (2018). doi:10.1101/332213
- 530 56. Wise, R., Greene, J., Büchel, C. & Scott, S. K. Brain regions involved in articulation.
- 531 *The Lancet* **353**, 1057–1061 (1999).
- 532 57. Houde, J. F., Nagarajan, S. S., Sekihara, K. & Merzenich, M. M. Modulation of the
- 533 Auditory Cortex during Speech: An MEG Study. *Journal of Cognitive Neuroscience* **14**,
- 534 1125–1138 (2002).
- 535 58. Belyk, M., Pfordresher, P. Q., Liotti, M. & Brown, S. The Neural Basis of Vocal Pitch

- 536 Imitation in Humans. *Journal of Cognitive Neuroscience* **28**, 621–635 (2016).
- 537 59. Behroozmand, R. *et al.* Sensory–motor networks involved in speech production and  
538 motor control: An fMRI study. *NeuroImage* **109**, 418–428 (2015).
- 539 60. Takaso, H., Eisner, F., Wise, R. J. S. & Scott, S. K. The Effect of Delayed Auditory  
540 Feedback on Activity in the Temporal Lobe While Speaking: A Positron Emission  
541 Tomography Study. *J Speech Lang Hear Res* **53**, 226–236 (2010).
- 542 61. Vaquero, L., Rodríguez-Fornells, A. & Reiterer, S. M. The Left, The Better: White-  
543 Matter Brain Integrity Predicts Foreign Language Imitation Ability. *Cerebral Cortex* **4**,  
544 2–12 (2016).
- 545 62. Kronfeld-Duenias, V., Amir, O., Ezrati-Vinacour, R., Civier, O. & Ben-Shachar, M.  
546 Dorsal and ventral language pathways in persistent developmental stuttering. *CORTEXX*  
547 **81**, 79–92 (2016).
- 548 63. Neef, N. E. *et al.* Left posterior-dorsal area 44 couples with parietal areas to promote  
549 speech fluency, while right area 44 activity promotes the stopping of motor responses.  
550 *NeuroImage* **142**, 628–644 (2016).
- 551 64. Chevillet, M. A., Jiang, X., Rauschecker, J. P. & Riesenhuber, M. Automatic Phoneme  
552 Category Selectivity in the Dorsal Auditory Stream. *Journal of Neuroscience* **33**, 5208–  
553 5215 (2013).
- 554 65. Markiewicz, C. J. & Bohland, J. W. Mapping the cortical representation of speech  
555 sounds in a syllable repetition task. *NeuroImage* **141**, 174–190 (2016).
- 556 66. Alho, J. *et al.* Early-latency categorical speech sound representations in the left inferior  
557 frontal gyrus. *NeuroImage* **129**, 214–223 (2016).
- 558 67. Du, Y., Buchsbaum, B. R., Grady, C. L. & Alain, C. Noise differentially impacts  
559 phoneme representations in the auditory and speech motor systems. *Proc Natl Acad Sci*  
560 *USA* **111**, 7126–7131 (2014).
- 561 68. Correia, J. M., Jansma, B. M. B. & Bonte, M. Decoding Articulatory Features from  
562 fMRI Responses in Dorsal Speech Regions. *Journal of Neuroscience* **35**, 15015–15025  
563 (2015).
- 564 69. Kanero, J., Imai, M., Okuda, J., Okada, H. & Matsuda, T. How Sound Symbolism Is  
565 Processed in the Brain: A Study on Japanese Mimetic Words. *PLoS ONE* **9**, e97905–8  
566 (2014).
- 567 70. Agnew, Z. K., McGettigan, C. & Scott, S. K. Discriminating between Auditory and  
568 Motor Cortical Responses to Speech and Nonspeech Mouth Sounds. *Journal of*  
569 *Cognitive Neuroscience* **23**, 4038–4047 (2011).
- 570 71. Krishnan, S. *et al.* Beatboxers and Guitarists Engage Sensorimotor Regions Selectively  
571 When Listening to the Instruments They can Play. *Cerebral Cortex* **28**, 4063–4079  
572 (2018).
- 573 72. Lewis, J. W., Talkington, W. J., Puce, A., Engel, L. R. & Frum, C. Cortical Networks  
574 Representing Object Categories and High-level Attributes of Familiar Real-world Action  
575 Sounds. *Journal of Cognitive Neuroscience* **23**, 2079–2101 (2011).
- 576 73. Engel, L. R., Frum, C., Puce, A., Walker, N. A. & Lewis, J. W. Different categories of  
577 living and non-living sound-sources activate distinct cortical networks. *NeuroImage* **47**,  
578 1778–1791 (2009).
- 579 74. Lewis, J. W., Brefczynski, J. A., Phinney, R. E., Janik, J. J. & DeYoe, E. A. Distinct  
580 Cortical Pathways for Processing Tool versus Animal Sounds. *Journal of Neuroscience*  
581 **25**, 5148–5158 (2005).

- 582 75. Repp, B. H. & Su, Y.-H. Sensorimotor synchronization: A review of recent research  
583 (2006–2012). *Psychonomic Bulletin & Review* **20**, 403–452 (2013).
- 584 76. Pfordresher, P. Q., Mantell, J. T., Brown, S., Zivadinov, R. & Cox, J. L. Brain responses  
585 to altered auditory feedback during musical keyboard production\_ An fMRI study. *Brain*  
586 *Research* **1556**, 28–37 (2014).
- 587 77. Gaver, W. W. What in the World Do We Hear?: An Ecological Approach to Auditory  
588 Event Perception. *Ecological Psychology* **5**, 1–29 (1993).
- 589 78. Warren, W. H. & Verbrugge, R. R. Auditory perception of breaking and bouncing  
590 events: A case study in ecological acoustics. *Journal of Experimental Psychology:*  
591 *Human Perception and Performance* **10**, 704–712 (1984).
- 592 79. Ortiz-Rios, M. *et al.* Widespread and Opponent fMRI Signals Represent Sound Location  
593 in Macaque Auditory Cortex. *Neuron* **93**, 971–983.e4 (2017).
- 594 80. Poirier, C. *et al.* Auditory motion-specific mechanisms in the primate brain. *PLoS Biol*  
595 **15**, e2001379 (2017).
- 596 81. Fiehler, K., Thaler, L., Schütz, I. & Meller, T. Neural Correlates of Human Echolocation  
597 of Path Direction During Walking. *Multisensory Research* **28**, 195–226 (2015).
- 598 82. Callan, A., Callan, D. E. & Ando, H. Neural correlates of sound externalization.  
599 *NeuroImage* **66**, 22–27 (2013).
- 600 83. Ceravolo, L., Frühholz, S. & Grandjean, D. Proximal vocal threat recruits the right  
601 voice-sensitive auditory cortex. *Social Cognitive and Affective Neuroscience* **11**, 793–  
602 802 (2016).
- 603 84. Ahveninen, J. *et al.* Evidence for distinct human auditory cortex regions for sound  
604 location versus identity processing. *Nature Communications* **4**, 615–19 (2013).
- 605 85. Zündorf, I. C., Lewald, J. & Karnath, H.-O. Testing the dual-pathway model for auditory  
606 processing in human cortex. *NeuroImage* **124**, 672–681 (2016).
- 607 86. Brungart, D. S. & Simpson, B. D. Within-ear and across-ear interference in a cocktail-  
608 party listening task. *The Journal of the Acoustical Society of America* **112**, 2985–2995  
609 (2002).
- 610 87. Phillips, D. P., Vigneault-MacLean, B. K., Boehnke, S. E. & Hall, S. E. Acoustic  
611 Hemifields in the Spatial Release from Masking of Speech by Noise. *Journal of the*  
612 *American Academy of Audiology* **14**, 518–524 (2003).
- 613 88. Mummery, C. J., Ashburner, J., Scott, S. K. & Wise, R. J. S. Functional neuroimaging of  
614 speech perception in six normal and two aphasic subjects. *The Journal of the Acoustical*  
615 *Society of America* **106**, 449–457 (1999).
- 616 89. Cohen, L., Jobert, A., Le Bihan, D. & Dehaene, S. Distinct unimodal and multimodal  
617 regions for word processing in the left temporal cortex. *NeuroImage* **23**, 1256–1270  
618 (2004).
- 619 90. Scott, S. K., McGettigan, C. & Eisner, F. A little more conversation, a little less action  
620 — candidate roles for the motor cortex in speech perception. *Nat Rev Neurosci* **10**, 295–  
621 302 (2009).
- 622 91. Zatorre, R. J., Belin, P. & Penhune, V. B. Structure and function of auditory cortex:  
623 music and speech. *Trends in Cognitive Sciences* **6**, 37–46 (2002).
- 624 92. Zatorre, R. J. & Belin, P. Spectral and Temporal Processing in Human Auditory Cortex.  
625 *Cerebral Cortex* **11**, 946–953 (2001).
- 626 93. Poeppel, D. The analysis of speech in different temporal integration windows: cerebral  
627 lateralization as ‘asymmetric sampling in time’. *Speech Communication* **41**, 245–255

- 628 (2003).
- 629 94. Scott, S. K., Blank, C. C., Rosen, S. & Wise, R. J. S. Identification of a pathway for  
630 intelligible speech in the left temporal lobe. *Brain* **123**, 2400–2406 (2000).
- 631 95. Wise, R. J. S. *et al.* Separate neural subsystems within 'Wernicke's area'. *Brain* **124**,  
632 83–95 (2001).
- 633 96. Winer, J. A., Miller, L. M., Lee, C. C. & Schreiner, C. E. Auditory thalamocortical  
634 transformation: structure and function. *Trends in Neurosciences* **28**, 255–263 (2005).
- 635 97. Bizley, J. K. in *Conn's Translational Neuroscience* 579–598 (Elsevier, 2017).  
636 doi:10.1016/B978-0-12-802381-5.00042-7
- 637 98. Chechik, G. *et al.* Reduction of Information Redundancy in the Ascending Auditory  
638 Pathway. *Neuron* **51**, 359–368 (2006).
- 639 99. Goldstein, J. L. Auditory Nonlinearity. *The Journal of the Acoustical Society of America*  
640 **41**, 676–699 (1967).
- 641 100. Fuchs, P. A., Glowatzki, E. & Moser, T. The afferent synapse of cochlear hair cells.  
642 *Current Opinion in Neurobiology* **13**, 452–458 (2003).
- 643 101. Harms, M. P. & Melcher, J. R. Sound Repetition Rate in the Human Auditory Pathway:  
644 Representations in the Waveshape and Amplitude of fMRI Activation. *Journal of*  
645 *Neurophysiology* **88**, 1433–1450 (2002).
- 646 102. Purcell, D. W., John, S. M., Schneider, B. A. & Picton, T. W. Human temporal auditory  
647 acuity as assessed by envelope following responses. *The Journal of the Acoustical*  
648 *Society of America* **116**, 3581–3593 (2004).
- 649 103. TAYLOR, W. R. & SMITH, R. G. The role of starburst amacrine cells in visual signal  
650 processing. *Vis Neurosci* **29**, 73–81 (2012).
- 651 104. Leff, A. P. *et al.* Impaired reading in patients with right hemianopia. *Annals of*  
652 *Neurology* **47**, 171–178 (2000).
- 653 105. Coslett, H. B., Brashear, H. R. & Heilman, K. M. Pure word deafness after bilateral  
654 primary auditory cortex infarcts. *Neurology* **34**, 347–352 (1984).
- 655 106. Ulanovsky, N., Las, L. & Nelken, I. Processing of low-probability sounds by cortical  
656 neurons. *Nat Neurosci* **6**, 391–398 (2003).
- 657 107. Polterovich, A., Jankowski, M. M. & Nelken, I. Deviance sensitivity in the auditory  
658 cortex of freely moving rats. *PLoS ONE* **13**, e0197678–22 (2018).
- 659 108. Yao, J. D., Bremen, P. & Middlebrooks, J. C. Emergence of Spatial Stream Segregation  
660 in the Ascending Auditory Pathway. *J. Neurosci.* **35**, 16199–16212 (2015).
- 661 109. Slutsky, D. A. & Recanzone, G. H. Temporal and spatial dependency of the  
662 ventriloquism effect. *Neuroreport* **12**, 7–10 (2001).
- 663 110. Chen, Y., Repp, B. H. & Patel, A. D. Spectral decomposition of variability in  
664 synchronization and continuation tapping: Comparisons between auditory and visual  
665 pacing and feedback conditions. *Human Movement Science* **21**, 515–532 (2002).
- 666 111. Kaas, J. H. & Hackett, T. A. Subdivisions of AuditoryCortex and Levels of Processing  
667 in Primates. *Audiology and Neurotology* **3**, 73–85 (1998).
- 668

669 **Acknowledgements**

670 During the preparation of this manuscript, K. J. was supported by an Early Career Fellowship from the Leverhulme Trust, and C. F. L. was supported by an FCT Investigator Grant from the  
671 Portuguese Foundation for Science and Technology (IF/00172/2015). The authors thank J. Rauschecker and R. Wise for immensely helpful discussions of the background to  
672 many of these studies.

673 **Author contributions**

674 The authors contributed equally to all aspects of the article.

675 **Publisher's note**

676 Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

677 **Reviewer information**

678 *Nature Reviews XXX* thanks [Referee#1 name], [Referee#2 name], and other anonymous reviewer(s), for their contribution to the peer review of this work.

679

680 **Competing Interests Statement**

681 The authors declare that there are no competing interests

682

683

684 **Box 1: The ascending and descending auditory pathways.**

685 Before sound is represented in the auditory cortex, it is first decomposed and undergoes extensive  
686 analysis in the ascending auditory pathway. For example, the spatial properties of sounds are  
687 known to be computed subcortically<sup>8,9,96</sup> and it is thus assumed that they do not need to be re-  
688 computed cortically. This subcortical processing is supplemented by further processing through  
689 cortico-thalamic loops, to enable auditory perception.

690

691 At the cochlea, the physical vibrations that give rise to the perception of sound are transduced into  
692 electrical signals. The cochlea encodes sound in a tonotopic form; that is, sounds of different  
693 frequencies are differentially represented. This tonotopic information is preserved within the  
694 auditory nerve and throughout the entire ascending auditory pathway into the core auditory cortical  
695 fields<sup>18</sup>. The auditory nerve fibres project from the cochlea to the cochlear nucleus (see the  
696 figure), where the auditory signal is decomposed into a number of parallel representations<sup>18</sup>.  
697 Divided into dorsal, anteroventral and posteroventral portions, the cochlear nucleus contains six  
698 principal cell types (as well as small cell and granule cell types), and mediates immensely complex  
699 processing of the auditory signal, which is only roughly characterised here. Each population of  
700 particular cochlear nucleus cell types receives input from across the whole tonotopic range and  
701 projects to a specific set of brain stem fields<sup>97</sup>. The anteroventral cochlear nucleus (AVCN) contains  
702 cells that respond to sounds with a very high level of temporal precision<sup>18</sup>. These project  
703 principally to the superior olivary nucleus and the trapezoidal body, which are important in  
704 computing the spatial location of sounds by comparing the inputs from the two ears, and thence to  
705 the inferior colliculus (IC)<sup>18</sup>. The posterodorsal cochlear nucleus (PVCN) contains cells which  
706 show responses to sound onsets, and repeated regular ('chopping') responses to sustained sounds:  
707 these PVCN cells display a broader range of frequency responses than those in AVCN<sup>18</sup>. The

708 dorsal cochlear nucleus (DCN) contains cells that display very complex frequency responses, such  
709 as highly specific frequency combination responses<sup>18</sup>. This may enable the identification of  
710 spectral ‘notches’, which are gaps in highly specific frequency ranges that are important for  
711 perceiving the spatial location of sound in the vertical plane. In addition to projecting to the  
712 superior olivary nucleus and trapezoidal body, the AVCN and PVCN both project to the lateral  
713 lemniscus and the IC directly<sup>18</sup>. The cochlear nucleus thus contributes to different sound  
714 processing pathways and contributes to the detection of a wide range of different informational  
715 aspects of incoming sounds, such as the spatial location of the source of the sound or the properties  
716 of the sound that can contribute to its identification (such as its pitch)<sup>97</sup>.

717 Further along the pathway, the IC is a critical relay station in the processing of sound:  
718 tonotopy is preserved and neurones are organised in sheets of cells that share common frequency  
719 responses. However, within a sheet neurons can vary in their responses to other aspects of sounds,  
720 such as their spatial location and amplitude characteristics<sup>97</sup>. Neural representations in the IC are  
721 less affected by a noisy and reverberant auditory environments than those of cochlear nucleus  
722 neurons, suggesting that the processing between these two regions makes the signal more robust,  
723 which may aid consistency in perceptual experience<sup>e.g. 98</sup>.

724 The IC projects to the auditory thalamus (including the medial geniculate nucleus, the  
725 medial pulvinar (PM) and the supragenulate nucleus of thalamus (Sg)/ limitans nucleus of  
726 thalamus (Lim) complex). The ventral medial geniculate nucleus (MGv) is, like the IC, organised  
727 tonotopically and is considered to be the main pathway to auditory cortex, though other thalamic  
728 nuclei project to auditory fields (Fig. 1a). The medial geniculate nucleus (MGm) receives auditory,  
729 visual somatosensory and vestibular inputs, and dorsal geniculate nuclei (MGad and MGpd) also  
730 receive auditory and somatosensory inputs: these cells tend to have fast, frequency specific  
731 responses to sounds<sup>97</sup>. These thalamic nuclei project to auditory core and surrounding auditory  
732 fields in the cortex (Figure 1a)<sup>13</sup>.

733 It is important to note that the primate auditory system does not faithfully transmit the auditory  
734 environment to the cortex. There is considerable loss of spectral detail at the cochlea, with a  
735 roughly logarithmic relationship between frequency and resolution meaning that the higher the  
736 frequency of the sound, the more compressed its resolution<sup>99</sup>. There is, however, reasonably good  
737 resolution of temporal detail at the cochlea, which is essential for the encoding of the interaural  
738 time differences that are used to compute spatial location of sounds<sup>100</sup>. At the inferior colliculus,

739 amplitude modulations with modulations rates slower than 200-300Hz (that is, those with a  
740 repetition rate around 3.3ms and longer) can be processed. However this temporal sensitivity  
741 reduces as the sounds are processed in the ascending auditory pathway <sup>101</sup>. For this reason  
742 perceptually humans are poor at detecting amplitude modulations with modulations rates that are  
743 faster than 50-60Hz (that is, those with a repetition length of than 16-20ms or longer)<sup>102</sup>.

744

745

746

## 747 **Box 2: Auditory and visual perception: differences**

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

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

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

787 The strong link between sounds, objects and actions may also underlie the robust finding  
788 that auditory sequences are far better than visual sequences for conveying a sense of rhythm<sup>110</sup>  
789 and auditory rhythms are much more salient than visual sequences. The link between sounds  
790 objects and actions also means that sounds can convey a great deal of information without  
791 necessarily being specifically recognised. A loud impact sound behind me will cause me to react,  
792 even if I cannot recognise exactly what hit what: it suggests that something large hit something  
793 else hard and whatever hit what, I might want to get out of the way.

794

795

## 796 **Fig. 1: Cortical and subcortical connectivity of the macaque auditory cortex**

797 a| A schematic representation of the connectivity between the auditory thalamus and core and  
798 belt auditory fields in the cortex. The connections broadly represent the proportion of the input to  
799 each auditory cortical field from the different thalamic nuclei. Connections that constitute a high  
800 proportion (over 40%) of the total connections to an auditory field from a given thalamic region

801 (based on the density of patterns of reciprocal staining) are indicated by solid lines; those that  
802 constitute a moderate proportion (between 10-40% of the total connections to an auditory field  
803 from a given thalamic region) are indicated with dashed lines, and low proportion (between 2%-  
804 10% of the total connections to an auditory field from a given thalamic region) are shown with  
805 dotted lines (for clarity, connections of 2% or fewer are not shown). There is a clear rostral–  
806 caudal distinction in thalamic connectivity. Moving rostrally, there is a general decline in the  
807 proportions of connections from the ventral division of the MGN (MGv) and increased  
808 proportions of inputs from other medial geniculate nuclei and other thalamic nuclei. Core areas  
809 A1 and R receive an overwhelming majority of their inputs from the ventral medial geniculate  
810 (MGv), while the more rostral RT area receives similar proportions of inputs from MGv and the  
811 postero-dorsal medial geniculate nucleus (MGpd). The rostromedial polar field (RTp) receives  
812 roughly similar proportions of its inputs from MGv, MGpd and the medial division of the medial  
813 geniculate nucleus (Mgm) as well as the suprageniculate nucleus of thalamus (Sg)/ limitans  
814 nucleus of thalamus (Lim) complex and the medial pulvinar (MP). Rostral superior temporal  
815 gyrus (STGr), belt and STG fields receive the majority of their thalamic inputs from the medial  
816 pulvinar, and a lower proportion from the Sg/Lim complex<sup>13</sup>. b| A schematic image illustrating  
817 the connectivity of different core auditory regions in the macaque cortex<sup>12,19,111</sup>. Dense feed  
818 forward, feedback, lateral and indeterminate connections (those for which the retrograde  
819 connectivity cell count was over 30) are represented with solid lines, whereas moderate  
820 feedforward and feedback connections (those for which cell count was between 15-29) are  
821 shown with dashed lines. The connectivity pattern shows a clear rostral-caudal difference: caudal  
822 core field A1 primarily connects to surrounding belt fields and to R, with more moderate  
823 connections to caudal belt and parabelt fields. R, on the other hand, connects to A1 and to rostral  
824 core field RT, with moderate connections to rostral and caudal belt and parabelt fields and RTp.  
825 RT connects to adjacent field RTp, and adjacent rostral belt fields. RTp has a distinctly different  
826 pattern of connectivity to temporal pole, rostral belt and parabelt fields, via lateral and  
827 indeterminate connections. This pattern of connectivity results in a recurrent and interactive  
828 network incorporating multiple parallel pathways with both direct and indirect connections<sup>12</sup>.  
829 AL, anterolateral belt; CL, caudolateral belt; CM, caudomedial belt; CPB, caudal parabelt;  
830 MGad, anterodorsal division of medial geniculate nucleus of thalamus (MGN); Mgm, medial  
831 division of MGN; MGpd, posterodorsal division of MGN; ML, middle lateral belt; MM, middle

832 medial belt; R, rostral auditory core field; RM, rostromedial belt; RT, rostral temporal core field;  
833 RTL, rostrotemporal-lateral belt; RTM, rostrotemporal-medial belt  
834 RPB, rostral parabelt; STS, superior temporal sulcus; TGdd, dysgranular part of the dorsal  
835 temporal pole; TGgd, granular part of the dorsal temporal pole; Tpt, temporo-parietal area.  
836 Part a is adapted, with permission from Scott et al 2017<sup>13</sup>, and part b is adapted from Scott et al,  
837 2015<sup>12</sup>.

838

839 **Fig. 2: Response properties of rostral and caudal auditory cortex.**

840 **a|** This shows the examples of ECoG responses to sentences categorized as ‘sustained’ and ‘onset’,  
841 based on machine learning classifications. Rostral fields show sustained responses, and caudal  
842 fields show transient responses associated with the onset of complex sequences. These distinctions  
843 are found bilaterally<sup>31</sup>. **b|** minimum response latencies (that is, the fastest responses to sound  
844 onsets) in rostral core field R (top) and caudal core field A1 (bottom). The median response in  
845 caudal A1 is faster (at 20ms) than that in rostral R (33ms)<sup>14</sup>. **c|** Neural responses to increasing  
846 rates of amplitude modulation in rostral core field R (top) and caudal core field A1 (bottom). Note  
847 that the responses saturate at a much lower amplitude modulation frequency in rostral field R than  
848 in caudal field A1, indicating that the responses in A1 can track amplitude changes at a much faster  
849 rate than can R<sup>14</sup>. Part a is adapted with permission from Ref.<sup>31</sup>. Parts b and c are adapted, with  
850 permission, from Ref<sup>14</sup>.

851