



ELSEVIER

SCIENCE @ DIRECT®

Auditory processing – speech, space and auditory objects

Sophie K Scott

There have been recent developments in our understanding of the auditory neuroscience of non-human primates that, to a certain extent, can be integrated with findings from human functional neuroimaging studies. This framework can be used to consider the cortical basis of complex sound processing in humans, including implications for speech perception, spatial auditory processing and auditory scene segregation.

Addresses

Institute of Cognitive Neuroscience, University College London, 17 Queen Square, London, WC1N 3AR, UK

Corresponding author: Scott, Sophie (sophie.scott@ucl.ac.uk)

Current Opinion in Neurobiology 2005, 15:1–5

This review comes from a themed issue on
Cognitive neuroscience
Edited by Angela D Friederici and Leslie G Ungerleider

0959-4388/\$ – see front matter
© 2005 Elsevier Ltd. All rights reserved.

DOI 10.1016/j.conb.2005.03.009

Introduction

In comparison with vision, auditory processing has traditionally been the poor relation of neuroscience. This is partly because of the technical difficulties involved in studying audition, both in recording from primate auditory areas and in stimulus selection and presentation, and because of the perceived dominance of vision — a dominance that neatly reverses if different tasks are used [1]. The processing demands of audition differ from those of vision in several important ways. First, sounds only have structure that evolves over time — in terms of both steady-state and changing aspects of the structure [2] — which potentially places different demands on the nature of auditory ‘memory’ [3]. Second, spatial information in sound needs to be reconstructed from two inputs (the binaural hearing system) [4], which has important consequences for the neurophysiology of auditory scene segregation. Whereas the primate auditory system solves many of the physical problems of auditory spectral and temporal structure and spatial organization subcortically [5], the information also needs to be represented cortically, and this probably contributes to acoustic scene segregation [6]. Third, sounds are generated by physical action — be it animate or inanimate. This means that information about actions is intimately associated with the nature of auditory representations, which is not

necessarily the case for static visual scenes [7[•]]. In this review, the cortical basis of audition in primates is considered with reference to auditory objects, scene segregation and actions. This also encompasses implications for speech perception.

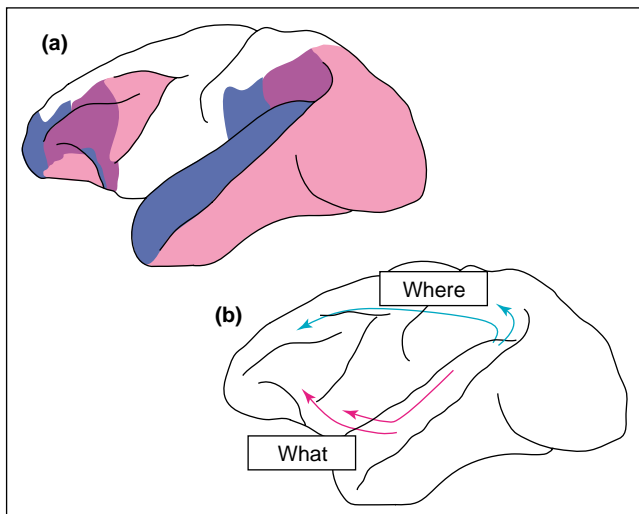
In a preceding review, Nelken [8] identified auditory cortex as having a role in the representation of auditory objects, rather than a role in the representation of invariant acoustic cues and features. This is an especially important suggestion because it has not been simple to establish the role of auditory cortex in hearing — ablating auditory cortex does not result in cortical deafness [9]. Rather, auditory cortex seems to be necessary for computing and representing complex acoustic properties of stimuli [10^{••},11]. Simplistic comparisons of the auditory cortex in humans with that in non-human primates remain controversial. However, in this review I assume that the general properties of non-human primate cortical processing are sufficiently similar [12[•]] to those in humans, and integrate findings from the two fields in an attempt to find commonalities.

From sounds to speech and space

Using positron emission tomography (PET) with monkeys, Poremba and co-workers [13] have demonstrated that extensive regions of primate cortex are responsive to acoustic stimulation (Figure 1). Importantly, these areas are located in frontal and temporal lobe regions adjacent to visually responsive cortex, with some areas of overlap. Within this widespread auditory system, there are now well-established patterns of connectivity from primary auditory cortex (PAC) (Figure 1). There are both hierarchically organized and parallel connections from PAC to belt and parabelt cortex, and projections from anterior and posterior auditory fields to premotor and prefrontal cortex. These connections have been expressly compared with those of the visual system, with respect to both the distinctive primate pattern of hierarchical organization of sensory cortex [14] and the partially distinct (although interacting) routes to anterior brain regions [15–17]. Similar to the situation in the visual system, there is a corresponding hierarchy of functional responses to acoustic stimulation; responses to pure tones can be observed in PAC, and responses to sounds with progressively greater signal bandwidths can be seen in lateral belt and parabelt. The response in the parabelt is organized cochleotopically in a rostral–caudal direction [18^{••}], with center frequency reversals that resemble those seen in core primary auditory cortical fields. There is also a functional specialization along the rostral–caudal dimension, with rostral parabelt regions showing an enhanced response to

2 Cognitive neuroscience

Figure 1

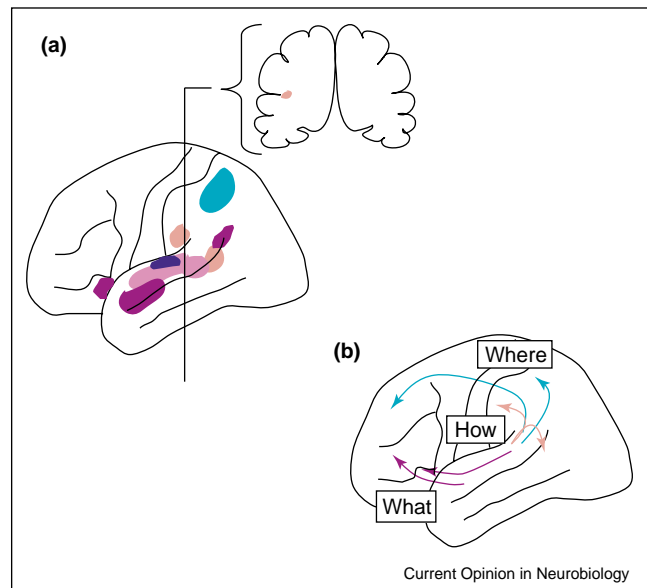


Auditory regions and streams in the primate brain. **(a)** The lateral surface of a macaque brain showing regions of visual (pink) and auditory (blue) responsiveness (adapted from Poremba *et al.* [13]). Multimodal responsiveness is shown in purple. **(b)** Two broad 'streams' of processing within the auditory system (adapted from Romanski *et al.* [17]).

conspecific vocalizations, and caudal parabelt regions showing greater sensitivity to the location in space of the calls [19]. The rostral–caudal distinction can also be seen in the response to more general properties of sounds: rostral lateral belt regions also respond preferentially to slower frequency modulated (FM) sweeps, whereas caudal lateral belt regions respond best to fast FM rates [20••].

This rostral–caudal distinction in function and anatomy has led to the proposal that the relatively distinct streams of processing can be fractionated along functional lines — an anterior or rostral 'what' pathway and a posterior or caudal 'where' pathway — and that this framework can be used to understand both lesion [21] and functional imaging studies [22••] in humans. Although aspects of the what–where distinction remain controversial [23,24], this is a framework that has generally gained support from human functional imaging studies. For example, posterior auditory or inferior parietal cortical responses are consistently seen across studies to sounds with spatial characteristics (e.g. moving sounds) [22••], and the planum temporale responds to speech that has a distinct free field 'outside the head' location (relative to 'inside the head') [25]. By contrast, the processing of linguistically relevant acoustic information is associated, in humans, with more anterior temporal lobe responses [22••]. This pattern of hierarchical processing within an anterior–posterior dimension has also been important in understanding the neural processing of speech (Figure 2; [26•]). In this model, the 'what' stream of processing, running lateral and anterior to PAC, is progressively more respon-

Figure 2



Functional responses to speech and candidate stream of processing in the human brain. **(a)** The lateral surface of the human brain, the coloured regions indicate broadly to which type of acoustic signal each temporal region (and associated parietal and frontal region) responds. Regions in blue show a specific response to language-specific phonological structure (Jacquemot *et al.* [27••]). Regions in lilac respond to stimuli with the phonetic cues and features of speech, whereas those in purple respond to intelligible speech (Scott *et al.* [31], Narain *et al.* [32]). Regions in pink respond to verbal short term memory and articulatory representations of speech (Wise *et al.* [39], Hickok *et al.* [38], Jacquemot *et al.* [27••]). Regions in green respond to auditory spatial tasks (Arnott *et al.* [22••]). **(b)** The putative directions of the 'what' 'where' and 'how' streams of processing in the human brain.

sive to intelligible speech along its length (running from posterior to anterior regions), regardless of whether or not the speech itself sounds human in origin. This general model has been elaborated on in more recent functional imaging studies; whereas speech-specific responses are not seen in PAC, a region of left superior temporal gyrus (STG) that is lateral to PAC (and possibly corresponding to the 'parabelt' in humans) has recently been shown to be sensitive to language-specific phonological structure (Figure 2; [27••]). This response is left lateralized, and might represent the start of the processing of speech information in the anterior 'what' pathway (Figure 2). The anterior direction of the processing of intelligible speech has also been observed using rapid event-related functional magnetic resonance imaging (fMRI) [28]. In more anterior fields, rostral to PAC, responses to both syntactic and semantic violations in sentences can be seen, implicating this anterior stream in the integration of lexical information in spoken language [29]. This study by Friederici *et al.* [29] also indicated that basal ganglia regions could be specifically associated with syntactic processing, evidence that the 'language system' as a whole is associated with regions beyond the temporal

lobes [30]. Further along the anterior stream, in the left anterior superior temporal sulcus (STS), responses are seen to intelligible speech [31,32], and this response is seen for both single words and sentences [33]. Thus, the anterior ‘what’ system in humans is important in the early stages of acoustic processing of speech. Responses to speech can be observed extending into frontal regions; ‘top down’ modulation of heard speech is associated with ventral prefrontal [34] and posterior premotor cortex activation [35,36]. This suggests a role for frontal auditory connections, and possibly motor representations, in spoken language processing. Indeed, some research has suggested that the pattern of responses in auditory cortex can be highly modulated by task related top-down processing [37].

In addition to a clear role for spatial processing of sound in posterior auditory fields in humans, there is evidence for at least two further kinds of speech-related auditory processing in posterior auditory fields, which might or might not form subsets of the same process. First, it has been suggested that aspects of verbal working memory are associated with left posterior STS [38,39] and supra-marginal gyrus [27]. This might relate to some of the issues of the nature of auditory memory — specifically, the need for transient representations that encode the temporal dimension [3]. In addition, medial posterior fields are activated during speech production [40] whether or not articulation is overt [39] or even specific to speech [38]. This implicates posterior auditory cortex in the guidance of the motor act of speech (and perhaps other motor acts), and might represent a sensory motor interface, involved in speech, that links perception and production. As mentioned in the Introduction, sounds convey information about the events that cause them, and a role for motor information has long been posited as a route for speech perception [41]. These posterior auditory–motor fields might, therefore, form part of the same system in which motor cortex [35] and left anterior insula [42] responses have been described in functional imaging studies, and contribute to a ‘how’ system in speech perception [24,26,33]. It is also striking that recordings from caudal medial auditory fields in primates have shown that they are responsive to touch — another potential link for sensory–motor integration [43,44]. The relationship among this putative ‘how’ pathway, transient auditory memory systems and the auditory ‘where’ pathways in humans requires further elaboration; they might all fall within a system that encodes spatial–motoric information generically, or they might form distinctly different subsystems.

Auditory objects, scenes and attention

How do the streams of auditory processing interact with auditory object processing and auditory scene analysis? It has been suggested that central auditory mechanisms are important for paying attention to auditory objects [45].

Single cell recordings from cat PAC have enabled investigators to identify the plasticity of response in PAC that is associated with the frequency of auditory objects [46]. In humans, fMRI has shown that primary auditory cortical fields are sensitive to the amplitude envelopes of sounds, and non primary auditory fields also show enhanced sensitivity to the onsets and offsets of sounds — phenomena associated with the structure of auditory events [47]. Evidence also shows that anterior auditory fields are important for the tracking of auditory streams of information [48]. Moving further from PAC in terms of synaptic distance, in a PET study Zatorre *et al.* [49] manipulated the acoustic cues of auditory objects to create the impression of multiple events. This revealed activation in right superior sulcus, anterior to primary auditory cortex, implicating the anterior ‘what’ stream in the representation of multiple auditory objects. We have also recently shown that there is extensive processing of an unattended speaker in lateral and anterior STG, suggesting that multiple complex auditory objects can be represented cortically, and thus providing a route for the semantic processing of ‘unattended’ speech [34]. Therefore, the ‘what’ stream of processing is apparently also implicated in the representation of, and perhaps the allocation of attention to, distinct auditory objects. How this interacts with posterior ‘where’ stream(s), which has also been associated with aspects of attention control of the auditory scene [6] and subcortical nuclei essential for the encoding of spatial cues, will be developed in further studies.

Returning to our discussion of non-human primates we look to the work of Poremba *et al.* [50], who have been investigating hemispheric lateralization for the processing of conspecific vocalizations using PET. They revealed an anterior superior temporal lobe response to meaningful vocalizations that was left lateralized. This response is strikingly similar to that seen to intelligible speech in human functional imaging studies [31]. Intriguingly, the asymmetric response was abolished following commissurotomy, suggesting that the diminished response to vocalizations in the right temporal pole was a result of activity on the left — perhaps an active suppression of the right by the left. Such suppression of the right hemisphere response has been noted in the right operculum in human studies of speech production [51]. It has proven difficult to account for hemispheric asymmetries in linguistic processing as a result of acoustic properties of the speech signal [33], although a recent study has suggested that such asymmetry derives from even simpler differences in auditory processing [52]. The nature of hemispheric differences in auditory and linguistic processing will be illuminated further by the characterization of such hemispheric interactions.

Conclusions

It is not fanciful to suggest that, as the most articulate primates, we have evolved a neural system optimized for

4 Cognitive neuroscience

aspects of speech perception and production, in contrast to other specializations (e.g. humans do not use hearing for hunting). Situating our understanding of speech, space and auditory objects in the context of the basic neuroanatomy of the primate auditory system is a strong position from which to elaborate on these early perceptual systems. I am optimistic that future work will develop the cortical and subcortical basis of the functional organization of human and non human hearing. I am also hopeful that the challenges of hemispheric asymmetries, interactions with attention and perception–production links will be addressed within a neuroanatomical framework.

Acknowledgements

SK Scott is funded by the Wellcome Trust, SRF GR074414MA. I would like to thank R Wise and S Rosen for helpful discussions on these topics.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- of special interest
 - of outstanding interest
1. Repp BH, Penel A: **Auditory dominance in temporal processing: new evidence from synchronization with simultaneous visual and auditory sequences.** *J Exp Psychol Hum Percept Perform* 2002, **28**:1085-1099.
 2. Handel S: **Space is to time as vision is to audition: seductive but misleading.** *J Exp Psychol Hum Percept Perform* 1988, **14**:315-317.
 3. Beaman CP, Morton J: **The separate but related origins of the recency effect and the modality effect in free recall.** *Cognition* 2000, **77**:B59-B65.
 4. Bregman AS: **Auditory scene analysis.** MIT press; 1990.
 5. Masterton RB: **Role of the central auditory system in hearing: the new direction.** *Trends Neurosci* 1992, **15**:280-285.
 6. Cusack R: **The intraparietal sulcus and perceptual organization.** *J Cogn Neurosci* in press.
 7. McAdams S, Chaigne A, Roussarie V: **The psychomechanics of simulated sound sources: material properties of impacted bars.** *J Acoust Soc Am* 2004, **115**:1306-1320.
This study uses synthesized sounds to vary the perceived material and geometric properties in a parametric manner. The results of multidimensional scaling studies show that listeners are sensitive to these mechanical properties, although the relationship between mechanical and perceptual scales was not linear.
 8. Nelken I: **Processing of complex stimuli and natural scenes in the auditory cortex.** *Curr Opin Neurobiol* 2004, **14**:474-480.
 9. Baran JA, Bothfeldt RW, Musiek FE: **Central auditory deficits associated with compromise of the primary auditory cortex.** *J Am Acad Audiol* 2004, **15**:106-116.
 10. Griffiths TD, Warren JD: **What is an auditory object?** *Nat Rev Neurosci* 2004, **5**:887-892.
This study sets out parameters for how auditory objects might be best described, and tries to reconcile the concept with ideas from visual processing and auditory scene segregation, in addition to models from auditory psychophysics.
 11. Griffiths TD, Warren JD, Scott SK, Nelken I, King AJ: **Cortical processing of complex sound: a way forward?** *Trends Neurosci* 2004, **27**:181-185.
 12. Wise RJ: **Language systems in normal and aphasic human subjects: functional imaging studies and inferences from animal studies.** *Br Med Bull* 2003, **65**:95-119.

In this study, the neurological models of language processing are explicitly contrasted with models of communication that are derived from non-

human primate models, with consequent implications for studies of patients and aphasia.

13. Poremba A, Saunders RC, Crane AM, Cook M, Sokoloff L, Mishkin M: **Functional mapping of the primate auditory system.** *Science* 2003, **299**:568-572.
14. Kaas JH, Hackett TA: **Subdivisions of auditory cortex and processing streams in primates.** *Proc Natl Acad Sci USA* 2000, **97**:11793-11799.
15. Romanski LM, Averbach BB, Diltz M: **Neural representation of vocalizations in the primate ventrolateral prefrontal cortex.** *J Neurophysiol* 2005, **93**:734-747.
16. Romanski LM, Goldman-Rakic PS: **An auditory domain in primate prefrontal cortex.** *Nat Neurosci* 2002, **5**:15-16.
17. Romanski LM, Tian B, Fritz J, Mishkin M, Goldman-Rakic PS, Rauschecker JP: **Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex.** *Nat Neurosci* 1999, **2**:1131-1136.
18. Rauschecker JP, Tian B: **Processing of band-passed noise in the lateral auditory belt cortex of the rhesus monkey.** *J Neurophysiol* 2004, **91**:2578-2589.
This study develops the authors' earlier work on lateral belt responses to band passed noise. It demonstrates that not only is there a medial–lateral increasing preference for wider bandwidths but there is an overall rostral–caudal response to center frequency, which mirrors the pattern of best tone frequencies in core regions.
19. Tian B, Reser D, Durham A, Kustov A, Rauschecker JP: **Functional specialization in rhesus monkey auditory cortex.** *Science* 2001, **292**:290-293.
20. Tian B, Rauschecker JP: **Processing of frequency-modulated sounds in the lateral auditory belt cortex of the rhesus monkey.** *J Neurophysiol* 2004, **92**:2993-3013.
This study uses the presentation of frequency modulated tones to probe differences in rostral–caudal lateral belt regions. The differentiation of the responses is evidence that previous findings of preference for conspecific vocalizations in rostral regions and of spatial locations in caudal regions are based on the more basic acoustic properties of the sounds.
21. Clarke S, Thiran AB: **Auditory neglect: what and where in auditory space.** *Cortex* 2004, **40**:291-300.
22. Arnott SR, Binns MA, Grady CL, Alain C: **Assessing the auditory dual-pathway model in humans.** *Neuroimage* 2004, **22**:401-408.
The authors present a thorough review of functional imaging studies of auditory processing in humans in a largely successful attempt to map these onto the domain specific what–where model of the non-human primate literature. This kind of meta-analysis has the advantage of overcoming some of the limitations of human studies (e.g. numbers of subjects, power of the analysis, scanner noise and use of explicit tasks).
23. Middlebrooks JC: **Auditory space processing: here, there or everywhere?** *Nat Neurosci* 2002, **5**:824-826.
24. Hickok G, Poeppel D: **Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language.** *Cognition* 2004, **92**:67-99.
25. Hunter MD, Griffiths TD, Farrow TF, Zheng Y, Wilkinson ID, Hegde N, Woods W, Spence SA, Woodruff PW: **A neural basis for the perception of voices in external auditory space.** *Brain* 2003, **126**:161-169.
26. Scott SK, Johnsrude IS: **The neuroanatomical and functional organization of speech perception.** *Trends Neurosci* 2003, **26**:100-107.
The authors attempt to use non-human primate models of auditory processing as a framework for human speech processing streams.
27. Jacquemot C, Pallier C, LeBihan D, Dehaene S, Dupoux E: **Phonological grammar shapes the auditory cortex: a functional magnetic resonance imaging study.** *J Neurosci* 2003, **23**:9541-9546.
Phonological structure refers in this study to the legality (or 'syntax') of phonological sequences (e.g. the sequence 'ebzu' is illegal in Japanese, because consonants must always be separated by a vowel), and also to the nature of phonetic contrasts — for example, vowel duration is not contrastive in French (thus 'ebuzu' and 'ebuuzu' sound like the same non-word in French). This study shows neural responses to changes in language-specific phonological structure, rather than simply to acoustic

change, which is an elegant way of probing the nature of perceptual responses to linguistic information. The result is also important because it shows a response 'early' in auditory cortex to aspects of the linguistic information in the structure of speech (which does not rely on the initial detection of phonemes).

28. Specht K, Reul J: **Functional segregation of the temporal lobes into highly differentiated subsystems for auditory perception: an auditory rapid event-related fMRI-task.** *Neuroimage* 2003, **20**:1944-1954.
29. Friederici AD, Ruschemeyer SA, Hahne A, Fiebach CJ: **The role of left inferior frontal and superior temporal cortex in sentence comprehension: localizing syntactic and semantic processes.** *Cereb Cortex* 2003, **13**:170-177.
30. Friederici AD: **Towards a neural basis of auditory sentence processing.** *Trends Cogn Sci* 2002, **6**:78-84.
31. Scott SK, Blank SC, Rosen S, Wise RJS: **Identification of a pathway for intelligible speech in the left temporal lobe.** *Brain* 2000, **123**:2400-2406.
32. Narain C, Scott SK, Wise RJS, Rosen S, Leff AP, Iversen SD, Matthews PM: **Defining a left-lateralised response specific to intelligible speech using fMRI.** *Cereb Cortex* 2003, **13**:1362-1368.
33. Scott SK, Wise RJS: **The functional neuroanatomy of prelexical processing of speech.** *Cognition* 2004, **92**:13-45.
34. Scott SK, Rosen S, Wickham L, Wise RJS: **A positron emission tomography study of the neural basis of informational and energetic masking effects in speech perception.** *J Acoust Soc Am* 2004, **115**:813-821.
35. Wilson SM, Saygin AP, Sereno MI, Iacoboni M: **Listening to speech activates motor areas involved in speech production.** *Nat Neurosci* 2004, **7**:701-702.
This study presents evidence that there are overlapping fields for speech perception and production in premotor and motor cortex.
36. Davis MH, Johnsrude IS: **Hierarchical processing in spoken language comprehension.** *J Neurosci* 2003, **23**:3423-3431.
37. Brechmann A, Scheich H: **Hemispheric shifts of sound representation in auditory cortex with conceptual listening.** *Cereb Cortex* in press.
38. Hickok G, Buchsbaum B, Humphries C, Muftuler T: **Auditory-motor interaction revealed by fMRI: speech, music, and working memory in area Spt.** *J Cogn Neurosci* 2003, **15**:673-682.
39. Wise RJS, Scott SK, Blank SC, Mummery CJ, Warburton E: **Identifying separate neural sub-systems within 'Wernicke's area'.** *Brain* 2001, **124**:83-95.
40. Blank SC, Scott SK, Murphy K, Warburton E, Wise RJ: **Speech production: Wernicke, Broca and beyond.** *Brain* 2002, **125**:1829-1838.
41. Liberman AM, Whalen DH: **On the relation of speech to language.** *Trends Cogn Sci* 2000, **4**:187-196.
42. Wise RJS, Greene J, Büchel C, Scott SK: **Brain systems for word perception and articulation.** *Lancet* 1999, **353**:1057-1061.
43. Fu KM, Johnston TA, Shah AS, Arnold L, Smiley J, Hackett TA, Garraghty PE, Schroeder CE: **Auditory cortical neurons respond to somatosensory stimulation.** *J Neurosci* 2003, **23**:7510-7515.
44. Schroeder CE, Smiley J, Fu KG, McGinnis T, O'Connell MN, Hackett TA: **Anatomical mechanisms and functional implications of multisensory convergence in early cortical processing.** *Int J Psychophysiol* 2003, **50**:5-17.
45. Darwin CJ: **Auditory Grouping.** *Trends Cogn Sci* 1997, **1**:327-333.
46. Ulanovsky N, Las L, Nelken I: **Processing of low-probability sounds by cortical neurons.** *Nat Neurosci* 2003, **6**:391-398.
47. Harms MP, Guinan JJ Jr, Sigalovsky IS, Melcher JR: **Short-term sound temporal envelope characteristics determine multisecond time-patterns of activity in human auditory cortex as shown by fMRI.** *J Neurophysiol* 2005, **95**:210-222.
This study builds on the authors' previous work to demonstrate sensitivity in cortical regions to the temporal characteristics of auditory sequences, and attempts to distinguish phasic and sustained responses across different rates.
48. Warren JD, Uppenkamp S, Patterson RD, Griffiths TD: **Separating pitch chroma and pitch height in the human brain.** *Proc Natl Acad Sci USA* 2003, **100**:10038-10042.
The authors present an elegant demonstration of different neural responses to changes in the perceived height of pitch, and changes in the chroma (i.e. place in scalar structure) of a pitch.
49. Zatorre RJ, Bouffard M, Belin P: **Sensitivity to auditory object features in human temporal neocortex.** *J Neurosci* 2004, **24**:3637-3642.
50. Poremba A, Malloy M, Saunders RC, Carson RE, Herscovitch P, Mishkin M: **Species-specific calls evoke asymmetric activity in the monkey's temporal poles.** *Nature* 2004, **427**:448-451.
This study is the first demonstration of a neural basis for a left hemisphere dominance in vocal processing in non-human primates, although there have been behavioral studies that suggested this might be the case.
51. Blank SC, Bird H, Turkheimer F, Wise RJ: **Speech production after stroke: the role of the right pars opercularis.** *Ann Neural* 2003, **54**:310-320.
52. Devlin JT, Raley J, Tunbridge E, Lanary K, Floyer-Lea A, Narain C, Cohen I, Behrens T, Jezzard P, Matthews PM, Moore DR: **Functional asymmetry for auditory processing in human primary auditory cortex.** *J Neurosci* 2003, **23**:11516-11522.