



Published in final edited form as:

Nat Neurosci. 2009 June ; 12(6): 718–724. doi:10.1038/nn.2331.

Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing

Josef P Rauschecker^{1,2} and Sophie K Scott³

¹Laboratory of Integrative Neuroscience and Cognition, Georgetown University Medical Center, Washington, DC, USA ²Centre of Excellence in Computational Complex Systems Research, Department of Biomedical Engineering and Computational Science, Helsinki University of Technology, Espoo, Finland ³Institute of Cognitive Neuroscience, University College London, London, UK

Abstract

Speech and language are considered uniquely human abilities: animals have communication systems, but they do not match human linguistic skills in terms of recursive structure and combinatorial power. Yet, in evolution, spoken language must have emerged from neural mechanisms at least partially available in animals. In this paper, we will demonstrate how our understanding of speech perception, one important facet of language, has profited from findings and theory in nonhuman primate studies. Chief among these are physiological and anatomical studies showing that primate auditory cortex, across species, shows patterns of hierarchical structure, topographic mapping and streams of functional processing. We will identify roles for different cortical areas in the perceptual processing of speech and review functional imaging work in humans that bears on our understanding of how the brain decodes and monitors speech. A new model connects structures in the temporal, frontal and parietal lobes linking speech perception and production.

Our understanding of speech processing has both benefited and suffered from developments in neuroscience. The basic brain areas important for speech perception and production were established in the nineteenth century, and although our conception of their exact anatomy and function has changed substantially, some of the findings of Broca¹ and Wernicke² still stand (Supplementary Discussion 1 and Supplementary Fig. 1 online). What has lagged behind is a good model of how the brain decodes spoken language and how speech perception and speech production are linked. For example, the frameworks for cortical processes and pathways have taken longer to form in audition than in vision, and animal models of language have severe limitations³. The evolution of speech and language are likely to have depended on neural systems available in other primate brains. In this paper, we will demonstrate how our understanding of speech perception, one important facet of language, has profited from work in nonhuman primate studies.

Streams and hierarchies in nonhuman primate auditory cortex

'What' and 'where' pathways in vision and audition

A decade ago, it was suggested that auditory cortical processing pathways are organized dually, similar to those in the visual cortex (Fig. 1)^{4,5}: that one main pathway projects from each of

Correspondence should be addressed to J.P.R. (rauschej@georgetown.edu) or S.K.S. (sophie.scott@ucl.ac.uk).

Note: Supplementary information is available on the Nature Neuroscience website.

the primary sensory areas into posterior parietal cortex, another pathway into anterior temporal cortex. As in the visual system⁶, the posterior parietal pathway was hypothesized to subserve spatial processing in audition while the temporal pathway subserved the identification of complex patterns or objects. Per the directions of their projections in the auditory system, these pathways were referred to as the postero-dorsal and antero-ventral streams, respectively.

Anatomical tract tracing studies in monkeys support separate anterior and posterior projection streams in auditory cortex^{7,8}. The long-range connections from the surrounding belt areas project from anterior belt directly to ventrolateral prefrontal cortex (PFC) and from the caudal (posterior) belt to dorsolateral PFC⁹. This latter finding provided evidence, on both anatomical and functional grounds^{10,11}, for ventral and dorsal processing streams within auditory cortex. Single-unit studies in the lateral belt areas of macaques provided more direct functional evidence for this dual processing scheme. Tian *et al.*¹² found that when species-specific communication sounds are presented in varying spatial locations, neurons in the antero-lateral belt (area AL) are more specific for the type of monkey call. By contrast, neurons in the caudo-lateral belt (area CL) are more responsive to spatial location than neurons in core or anterior belt. This result indicates that ‘what’ processing dissociates from ‘where’ processing in rhesus monkey auditory cortex.

The dual-stream hypothesis has found support from other studies^{13,14}. Recanzone and co-workers¹⁵ found a tighter correlation of neuronal activity and sound localization in caudal belt, supporting a posterior ‘where’ stream. Lewis and Van Essen¹⁶ described a direct auditory projection from the posterior superior temporal (pST) region to the ventral inferior parietal (VIP) area in the posterior parietal cortex of the monkey. Single-unit as well as imaging studies in monkeys also reveal functional specialization^{17–21}.

Functional magnetic resonance imaging in nonhuman primates identified, first, tonotopic maps on the superior temporal plane and gyrus²² and, then, a ‘voice region’ in the anterior part of the superior temporal gyrus²³, a voice region that projects further to the anterior superior temporal sulcus and ventrolateral PFC²⁴. Reversible cortical inactivation (using cortical cooling) in cat auditory cortex²⁵ found that inactivating anterior areas leads to a deterioration of auditory pattern discrimination, whereas inactivating posterior areas impairs spatial discrimination. These studies corroborate the notion that an antero-ventral processing stream forms the substrate for the recognition of auditory objects, including communication sounds, whereas a postero-dorsal stream includes spatial perception as at least one of its functions.

Hierarchical organization in the cerebral cortex combines elements of serial as well as parallel processing: ‘lower’ cortical areas with simpler receptive-field organization, such as sensory core areas, project to ‘higher’ areas with increasingly complex response properties, such as belt, parabelt and PFC regions. These complex properties are generated by convergence and summation (Box 1 and Fig. 2). Parallel processing principles in hierarchical organization are evident in that specialized cortical areas (‘maps’) with related functions (corresponding to submodalities or modules) are bundled into parallel processing ‘streams’. Furthermore, highly interconnected neural networks, dynamically modulated by different task demands, may also exist within hierarchical processing structures, and well known feedback connections are sometimes not sufficiently accounted for in hierarchical models.

‘What’ and ‘how’ pathways and the perception–action cycle

In addition to the ‘what/where’ model in vision⁶, Goodale and Milner²⁶ proposed that two pathways subserve behaviors related to perception and action. The auditory ventral pathway role in perception is largely consistent with a ‘what’ pathway, whereas the dorsal pathway takes on a sensorimotor role involved in action (‘how’), including spatial analysis. Fuster²⁷ advocates a similar distinction with regard to PFC and unites the two pathways into a perception–action

cycle. We argue here that the ‘what/where’ and ‘perception/action’ theories differ mainly in emphasis.

Dual processing streams in the auditory cortex of humans

The concepts of auditory streams of processing can be a powerful framework for understanding functional imaging studies of speech perception^{28,29} and for understanding aphasic stroke³. Human studies also confirm the role of the postero-dorsal stream in the perception of auditory space and motion (see refs. 30 and 14 for review). But do more than two processing streams exist³¹ (Fig. 3)? The posterior superior temporal gyrus and inferior parietal cortex have long been implicated in the processing of speech and language, and ignoring these reports (Supplementary Discussion 1) and assigning an exclusively spatial function to the postero-dorsal auditory stream would be unwise. It is therefore essential to discuss how the planum temporale, the temporoparietal junction and the inferior parietal cortex are involved in speech and language, and whether we can assign a common computational function to the postero-dorsal stream that encompasses both spatial and language functions.

Antero-ventral stream for auditory object and speech perception Hierarchical organization

A meta-analysis of imaging studies of speech processing³² reports an antero-lateral gradient along which the complexity of preferred stimuli increases, from tones and noise bursts to words and sentences. As in nonhuman primates, frequency responses show tonotopy, while core regions responding to tones are surrounded by belt areas preferring band-pass noise bursts³³. Using high-field scanners, multiple tonotopic fields³⁴ and multiple processing levels (core, belt and parabelt)³⁵ can be identified in human auditory cortex.

Auditory object identification

This sort of hierarchical organization in the antero-ventral auditory pathway of humans is important in auditory pattern recognition and object identification. As in animal models, preferred features of lower-order neurons combine to create selectivity for increasingly complex sounds^{36,37}, and regions can be seen that are specialized in different auditory object classes (A.M. Leaver and J.P.R., unpublished data)^{38,39}. Developments in how we conceive the structure of auditory objects^{40,41} will help extend these kinds of investigations. Like their visual counterparts, auditory objects coexist based on many attributes, such as timbre, pitch and loudness, that give each its distinctive perceptual identity⁴¹.

Speech and voice perception

Within speech perception, there is evidence that speech sounds are hierarchically encoded, as the anterior superior temporal cortex responds as a function of speech intelligibility, and not stimulus complexity alone^{42–44}. Similarly, Liebenthal *et al.*⁴⁵ and Obleser *et al.*⁴⁶ showed that the left middle and anterior superior temporal sulcus is more responsive to consonant–vowel syllables than auditory baselines. Thus, regions within the ‘what’ stream show the first clear responses to abstract, linguistic information in speech. Within these speech-specific regions of anterior superior temporal cortex, there may be subregions selective for particular speech-sound classes, such as vowels^{38,46}, raising the possibility that phonetic maps have some anatomical implementation in anterior temporal lobe areas.

Activity related to speaker recognition also exists in antero-lateral temporal lobe areas³⁹, sometimes extending into midtemporal regions as well. These human voice regions may be homologous, according to crude topological criteria, to monkey areas²³ mentioned above. This human ‘voice area’ in the anterior auditory fields seems to process detailed spectral properties of talkers⁴⁷. Notably, speech perception and voice discrimination dissociate clinically,

suggesting that the two are supported by different systems within the anterior and middle temporal lobes.

Invariance and categorization

An important problem in the task of speech perception is that of invariance against distortions in the scale of frequency (for example, pitch changes; Fig. 4a) or time (for example, compressions). For example, noise-vocoded speech, which simulates aspects of speech after cochlear implantation, is quite coarse in its spectro-temporal representation⁴⁸ (Fig. 4b); it is, however, readily intelligible after a brief training session. Perceptual invariance is also important in the perception of normal speech, as the ‘same’ phoneme can be acoustically very different (owing to coarticulation) and still be identified as the same sound⁴⁹: the sound /s/ is different at the start of “sue” than at the start of “see,” but remains an /s/.

Box 1

Hierarchical processing and combination sensitivity

Functional specialization and streams of processing are central to theories of hierarchical organization. Cortical specialization is generated by specificity at the level of single neurons. Their complex response properties are in turn generated by convergence from lower-order neurons and nonlinear summation—‘combination sensitivity’ (Fig. 2). Discovered originally in bats⁹⁷ and songbirds⁹⁸, combination sensitivity has been demonstrated in nonhuman primates as well¹⁷. It is a fundamental mechanism for generating highly selective neurons (or small networks), as required for speech perception. Such higher-order specificity is generated by combining input from lower-level neurons specific to relatively simple features. Thus combination sensitivity is an example of hierarchical processing at the cellular level. Because it necessitates single-neuron recording techniques, it can only be explored in animal models. Therefore, it is an example of how animal research in general has led to an understanding of speech perception at the cellular level and how animal models will remain necessary to obtain a complete understanding of the neural mechanisms of speech perception that goes beyond localization of function.

These examples of perceptual constancy are computationally difficult to solve. This ability to deal with invariance problems is not unique to speech or audition; it is a hallmark of all higher cortical perceptual systems. The structural and functional organization of the anterior-ventral streams in both the visual and auditory systems could illustrate how the cerebral cortex solves this problem. For example, it has been suggested that visual categories are formed in the lateral PFC⁵⁰, which receives input from higher-order object representations in the anterior temporal lobe¹⁰. In audition, using species-specific communication sounds, Romanski *et al.*⁵¹ found clusters of neurons in the macaque ventrolateral PFC encoding similar complex calls, and category-specific cells encoding single semantic categories have also been reported⁵². In humans, rapid adaptation studies with functional MRI in the visual system have recently led to similar conclusions⁵³. The invariance problem in speech perception may be solved in the inferior frontal cortex, or by interactions between inferior frontal and anterior superior temporal cortex.

Hemispheric asymmetry

Speech perception and production are left-lateralized in the human brain (for example, refs. 3, 42,54), and there is considerable interest in the neural basis of this (for example, ref. 55).

Hemispheric specialization is an important feature of the human brain, particularly in relation to speech and spatial processing. It remains to be seen to what extent animal models can contribute to our understanding of these asymmetries.

Postero-dorsal auditory stream for space and speech

Evidence for a postero-dorsal stream in auditory spatial processing is just as strong, if not stronger, in the human as in nonhuman primates. Stroke studies as well as modern neuroimaging have shown that spatial processing in the temporo-parietal cortex is often right-lateralized in humans, contralateral to language. Generally, spatial neglect is more frequent and severe after damage to the right hemisphere. We cannot discuss all pertinent results in this focus paper, but we refer the reader to other reviews (refs. 14,30; see also Supplementary Discussion 2 online).

The pST region (or planum temporale) in humans (and the dorsal stream emanating from it) has classically been assigned a role in speech perception⁵⁶. This contradicts the evidence for a spatial role for pST, as well as a more anterior location for speech sound decoding, as discussed above (see also Supplementary Discussions 1 and 2). One unifying view is that the planum temporale is generally involved in the processing of spectro-temporally complex sounds⁴⁶, which includes music processing⁵⁷. According to this view, the planum temporale operates as a ‘computational hub’⁵⁸.

The inferior parietal lobule (IPL), particularly the angular and supramarginal gyri (Brodmann areas 39 and 40), has also been linked to linguistic functions⁵⁹, such as the ‘phonological-articulatory loop’⁶⁰. Functional imaging has confirmed this role, though activity varies with working memory task load^{61,62}. However, the IPL does not seem to be driven by acoustic processing of speech: the angular gyrus (together with extensive prefrontal activation) is recruited when higher-order linguistic factors improve speech comprehension⁶³, rather than by acoustic influences on intelligibility. Thus the parietal cortex is associated with more domain-general, linguistic factors in speech comprehension, rather than acoustic or phonetic processing.

Multisensory reference frames in the postero-dorsal stream

There is now neurophysiological evidence that auditory caudal belt areas are not solely responsive to auditory input but show multimodal responses^{64,65}: both caudal medial and lateral belt fields receive input from somatosensory and multisensory cortex. Thus any spatial transformations conducted in the postero-dorsal stream may be based on a multisensory reference frame^{66,67}.

These multisensory responses in caudal auditory areas may underlie some functional specificity in humans. Several studies of silent articulation⁶⁸ and nonspeech auditory stimuli⁶⁹ find activation in a posterior medial planum temporale region, within the postero-dorsal stream. The medial planum temporale in man⁷⁰ has been associated with the representation of templates for “doable” articulations and sounds (not limited to speech sounds). This approach can be compared to the “affordance” model of Gibson^{71,72}, in which objects and events are described in terms of action possibilities. Such a sensorimotor role for the dorsal stream is consistent with the notion of an “action” stream in vision²⁶. The concept can be extended to auditory-motor transformations in verbal working memory tasks^{73,74} that involve articulatory representations^{60,75}. The postero-medial planum temporale area has also been identified as a key node for the control of speech production⁵⁴, as it shows a response to somatosensory input from articulators.

Speech perception–production links

There is considerable neural convergence between speech perception and production systems. For example, the postero-medial planum temporale area described in the previous section is an auditory area important in the motor act of articulation. Conversely, real or imagined speech

sounds and music result in activation within premotor areas important in overt production of speech⁷⁶ and music^{77,78}. Within auditory areas, monkey studies have shown that auditory neurons are suppressed during the monkey's own vocalizations^{79,80}. This finding is consistent with results from humans indicating that superior temporal areas are suppressed during speech production^{81,82} and that the response to one's own voice is always less than the response to someone else's.

At one level these findings may simply reflect the ways that sensory responses to actions caused by oneself are always differently processed from those caused by the actions of others⁸³, and this may support mechanisms important in differentiating between one's own voice and the voices of others. In primate studies, however, auditory neurons that are suppressed during vocalizations are often more activated if the sound of the vocalizations is distorted⁸⁰. This might indicate a specific role for these auditory responses in the comparison of feedforward and feedback information from the motor and auditory system during speech production⁸⁴. Distorting speech production in real time reveals enhanced activation in bilateral (posterior temporal) auditory fields to distorted feedback⁸⁵. New work using high-resolution diffusion tensor imaging in humans has revealed that there are direct projections from the pars opercularis of Broca's area (Brodmann area 44) to the IPL⁸⁶, in addition to the ones from ventral premotor cortex⁸⁷. With the known connections between parietal cortex and posterior auditory fields, this could form the basis for feed-forward connections between speech production areas and posterior temporal auditory areas (Fig. 5).

Common computational function of the postero-dorsal stream

The dual-stream processing model in audition^{4,5} has been a useful construct in hearing research, perceptual physiology and, in particular, psycholinguistics, where it has spawned several further models^{73,74} that have tried to accommodate specific results from this field. The role of a ventral stream in hierarchical processing of objects, as in the visual system, is now widely accepted. Specifically for speech, anterior regions of the superior temporal cortex respond to native speech sounds and intelligible speech, and these sounds are mapped along phonological parameter domains. By contrast, early posterior regions in and around the planum temporale are involved in the processing of many different types of complex sound. Later posterior regions participate in the processing of auditory space and motion but seem to integrate input from several other modalities as well.

Although evidence is strong for the role of the dorsal pathway (including pST) in space processing, the dorsal pathway needs to accommodate speech and language functions as well. Spatial transformations may be one example of fast adaptations used by 'internal models' or 'emulators', as first developed in motor control theory. Within these models, 'forward models' (predictors) can be used to predict the consequences of actions, whereas 'inverse models' (controllers) determine the motor commands required to produce a desired outcome⁸⁸. More recently, forward models have been used to describe the predictive nature of perception and imagery⁸⁹. The IPL could provide an ideal interface, where feed-forward signals from motor preparatory networks in the inferior frontal cortex and premotor cortex (PMC) can be matched with feedback signals from sensory areas⁷².

In speech perception and production, projections from articulatory networks in Broca's area and PMC to the IPL and pST interact with signals from auditory cortex (Fig. 5). The feed-forward projection from Brodmann area 44 (and ventral PMC) may provide an efference copy in the classical sense of von Holst and Mittelstaedt⁹⁰, informing the sensory system of motor articulations that are about to happen. This occurs in anticipation of a motor signal if the behavior is enacted, or as imagery if it is not. The activity arriving in the IPL and pST from frontal areas anticipates the sensory consequences of action. The feedback signal coming to

the IPL from pST, conversely, could be considered an “afference copy”⁹¹ with relatively short latencies and high temporal precision⁹²—a sparse but fast primal sketch of ongoing sensory events⁹³ that are compared with the predictive motor signal in the IPL at every instance.

‘Internal model’ structures in the brain are generally thought to enable smooth sequential motor behaviors, from visuospatial reaching to articulation of speech. The goal of these models is to minimize the resulting error signal through adaptive mechanisms. At the same time, these motor behaviors also support aspects of perception, such as stabilization of the retinal image and disambiguation of phonological information, thus switching between forward and inverse modes. As Indefrey and Levelt⁹⁴ point out, spoken language “constantly operates a dual system, perceiving and producing utterances. These systems not only alternate, but in many cases they partially or wholly operate in concert.” What is more, both spatial processing and real-time speech processing make use of the same internal model structures.

In summary, our new model of the auditory cortical pathways builds on the previous model of dual processing pathways for object identification and spatial analysis^{5,6}, but integrates the spatial (dorsal) pathway with findings from speech and music processing as well. The model is based on neuroanatomical data from nonhuman primates, operating under the assumption that mechanisms of speech and language in humans have built on structures available in other primates. Finally, our new model extends beyond speed processing⁷⁴ and applies in a very general sense to both vision and audition, in its relationship with previous models of perception and action^{26,27}.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We wish to thank D. Klemm for help with graphic design and T. Tan for help with editing. The work was supported by grants from the US National Institutes of Health (R01NS52494) and the US National Science Foundation (BCS-0519127 and PIRE-OISE-0730255) to J.P.R., and by Wellcome Trust Grant WT074414MA to S.K.S.

References

1. Broca P. Remarques sur le siège de la faculté du langage articulé: suivies d'une observation d'aphémie (perte de la parole). *Bull Soc Anat Paris* 1861;6:330–357.
2. Wernicke, C. *Der aphasische Symptomencomplex: Eine psychologische Studie auf anatomischer Basis.* Cohn & Weigert, Breslau; Germany: 1874.
3. 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. [PubMed: 12697619]
4. Rauschecker JP. Cortical processing of complex sounds. *Curr Opin Neurobiol* 1998;8:516–521. [PubMed: 9751652]
5. Rauschecker JP, Tian B. Mechanisms and streams for processing of “what” and “where” in auditory cortex. *Proc Natl Acad Sci USA* 2000;97:11800–11806. [PubMed: 11050212]
6. Mishkin M, Ungerleider LG, Macko KA. Object vision and spatial vision: two cortical pathways. *Trends Neurosci* 1983;6:414–417.
7. Kaas JH, Hackett TA. Subdivisions of auditory cortex and processing streams in primates. *Proc Natl Acad Sci USA* 2000;97:11793–11799. [PubMed: 11050211]
8. Hackett TA, Stepniewska I, Kaas JH. Subdivisions of auditory cortex and ipsilateral cortical connections of the parabelt auditory cortex in macaque monkeys. *J Comp Neurol* 1998;394:475–495. [PubMed: 9590556]
9. Romanski LM, et al. Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nat Neurosci* 1999;2:1131–1136. [PubMed: 10570492]

10. Goldman-Rakic PS. The prefrontal landscape: implications of functional architecture for understanding human mentation and the central executive. *Phil Trans R Soc Lond B* 1996;351:1445–1453. [PubMed: 8941956]
11. Petrides M. Lateral prefrontal cortex: architectonic and functional organization. *Phil Trans R Soc Lond B* 2005;360:781–795. [PubMed: 15937012]
12. Tian B, Reser D, Durham A, Kustov A, Rauschecker JP. Functional specialization in rhesus monkey auditory cortex. *Science* 2001;292:290–293. [PubMed: 11303104]
13. Schreiner CE, Winer JA. Auditory cortex mapmaking: principles, projections, and plasticity. *Neuron* 2007;56:356–365. [PubMed: 17964251]
14. Recanzone GH, Sutter ML. The biological basis of audition. *Annu Rev Psychol* 2008;59:119–142. [PubMed: 17678445]
15. Recanzone GH, Guard DC, Phan ML, Su TK. Correlation between the activity of single auditory cortical neurons and sound-localization behavior in the macaque monkey. *J Neurophysiol* 2000;83:2723–2739. [PubMed: 10805672]
16. Lewis JW, Van Essen DC. Corticocortical connections of visual, sensorimotor, and multimodal processing areas in the parietal lobe of the macaque monkey. *J Comp Neurol* 2000;428:112–137. [PubMed: 11058227]
17. Rauschecker JP, Tian B, Hauser M. Processing of complex sounds in the macaque nonprimary auditory cortex. *Science* 1995;268:111–114. [PubMed: 7701330]
18. Poremba A, et al. Functional mapping of the primate auditory system. *Science* 2003;299:568–572. [PubMed: 12543977]
19. 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. [PubMed: 15486426]
20. 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. [PubMed: 15136602]
21. Bendor D, Wang X. The neuronal representation of pitch in primate auditory cortex. *Nature* 2005;436:1161–1165. [PubMed: 16121182]
22. Petkov CI, Kayser C, Augath M, Logothetis NK. Functional imaging reveals numerous fields in the monkey auditory cortex. *PLoS Biol* 2006;4:e215. [PubMed: 16774452]
23. Petkov CI, et al. A voice region in the monkey brain. *Nat Neurosci* 2008;11:367–374. [PubMed: 18264095]
24. Kikuchi Y, et al. Voice region connectivity in the monkey assessed with microstimulation and functional imaging. *Soc Neurosci Abstr* 2008;850.2
25. Lomber SG, Malhotra S. Double dissociation of ‘what’ and ‘where’ processing in auditory cortex. *Nat Neurosci* 2008;11:609–616. [PubMed: 18408717]
26. Goodale MA, Milner AD. Separate visual pathways for perception and action. *Trends Neurosci* 1992;15:20–25. [PubMed: 1374953]
27. Fuster, J. *The Prefrontal Cortex*. Academic; London: 2008.
28. Scott SK, Johnsrude IS. The neuroanatomical and functional organization of speech perception. *Trends Neurosci* 2003;26:100–107. [PubMed: 12536133]
29. Scott SK. Auditory processing—speech, space and auditory objects. *Curr Opin Neurobiol* 2005;15:197–201. [PubMed: 15831402]
30. Rauschecker, JP. Cortical processing of auditory space: pathways and plasticity. In: Mast, F.; Jäncke, L., editors. *Spatial Processing in Navigation, Imagery and Perception*. Springer; New York: 2007. p. 389–410.
31. Kaas JH, Hackett TA. ‘What’ and ‘where’ processing in auditory cortex. *Nat Neurosci* 1999;2:1045–1047. [PubMed: 10570476]
32. Binder JR, et al. Human temporal lobe activation by speech and nonspeech sounds. *Cereb Cortex* 2000;10:512–528. [PubMed: 10847601]
33. Wessinger CM, et al. Hierarchical organization of the human auditory cortex revealed by functional magnetic resonance imaging. *J Cogn Neurosci* 2001;13:1–7. [PubMed: 11224904]
34. Formisano E, et al. Mirror-symmetric tonotopic maps in human primary auditory cortex. *Neuron* 2003;40:859–869. [PubMed: 14622588]

35. Chevillet M, Riesenhuber M, Rauschecker JP. Functional localization of the auditory “what” stream hierarchy. *Soc Neurosci Abstr* 2007;174.9
36. Zatorre RJ, Bouffard M, Belin P. Sensitivity to auditory object features in human temporal neocortex. *J Neurosci* 2004;24:3637–3642. [PubMed: 15071112]
37. Patterson RD, Uppenkamp S, Johnsrude IS, Griffiths TD. The processing of temporal pitch and melody information in auditory cortex. *Neuron* 2002;36:767–776. [PubMed: 12441063]
38. Obleser J, et al. Vowel sound extraction in anterior superior temporal cortex. *Hum Brain Mapp* 2006;27:562–571. [PubMed: 16281283]
39. Kumar S, Stephan KE, Warren JD, Friston KJ, Griffiths TD. Hierarchical processing of auditory objects in humans. *PLOS Comput Biol* 2007;3:e100. [PubMed: 17542641]
40. Shamma S. On the emergence and awareness of auditory objects. *PLoS Biol* 2008;6:e155. [PubMed: 18578570]
41. Scott SK, Blank CC, Rosen S, Wise RJ. Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* 2000;123:2400–2406. [PubMed: 11099443]
42. Narain C, et al. Defining a left-lateralized response specific to intelligible speech using fMRI. *Cereb Cortex* 2003;13:1362–1368. [PubMed: 14615301]
43. Scott SK, Rosen S, Lang H, Wise RJ. Neural correlates of intelligibility in speech investigated with noise vocoded speech—a positron emission tomography study. *J Acoust Soc Am* 2006;120:1075–1083. [PubMed: 16938993]
44. Liebenthal E, Binder JR, Spitzer SM, Possing ET, Medler DA. Neural substrates of phonemic perception. *Cereb Cortex* 2005;15:1621–1631. [PubMed: 15703256]
45. Obleser J, Zimmermann J, Van Meter J, Rauschecker JP. Multiple stages of auditory speech perception reflected in event-related fMRI. *Cereb Cortex* 2007;17:2251–2257. [PubMed: 17150986]
46. Belin P, Zatorre RJ. Adaptation to speaker's voice in right anterior temporal lobe. *Neuroreport* 2003;14:2105–2109. [PubMed: 14600506]
47. Warren JD, Scott SK, Price CJ, Griffiths TD. Human brain mechanisms for the early analysis of voices. *Neuroimage* 2006;31:1389–1397. [PubMed: 16540351]
48. Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M. Speech recognition with primarily temporal cues. *Science* 1995;270:303–304. [PubMed: 7569981]
49. Bailey PJ, Summerfield Q. Information in speech: observations on the perception of [s]-stop clusters. *J Exp Psychol Hum Percept Perform* 1980;6:536–563. [PubMed: 6447767]
50. Freedman DJ, Riesenhuber M, Poggio T, Miller EK. Categorical representation of visual stimuli in the primate prefrontal cortex. *Science* 2001;291:312–316. [PubMed: 11209083]
51. Romanski LM, Averbach BB, Diltz M. Neural representation of vocalizations in the primate ventrolateral prefrontal cortex. *J Neurophysiol* 2005;93:734–747. [PubMed: 15371495]
52. Russ BE, Ackelson AL, Baker AE, Cohen YE. Coding of auditory-stimulus identity in the auditory non-spatial processing stream. *J Neurophysiol* 2008;99:87–95. [PubMed: 18003874]
53. Jiang X, et al. Categorization training results in shape- and category-selective human neural plasticity. *Neuron* 2007;53:891–903. [PubMed: 1735923]
54. Dhanjal NS, Handunnetthi L, Patel MC, Wise RJ. Perceptual systems controlling speech production. *J Neurosci* 2008;28:9969–9975. [PubMed: 18829954]
55. Boemio A, Fromm S, Braun A, Poeppel D. Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nat Neurosci* 2005;8:389–395. [PubMed: 15723061]
56. Geschwind N. Disconnexion syndromes in animals and man. *Brain* 1965;88:237–294. [PubMed: 5318481]
57. Hyde KL, Peretz I, Zatorre RJ. Evidence for the role of the right auditory cortex in fine pitch resolution. *Neuropsychologia* 2008;46:632–639. [PubMed: 17959204]
58. Griffiths TD, Warren JD. The planum temporale as a computational hub. *Trends Neurosci* 2002;25:348–353. [PubMed: 12079762]
59. Caplan D, Rochon E, Waters GS. Articulatory and phonological determinants of word length effects in span tasks. *Q J Exp Psychol A* 1992;45:177–192. [PubMed: 1410554]
60. Baddeley A, Lewis V, Vallar G. Exploring the articulatory loop. *Q J Exp Psychol A* 1984;36:233–252.

61. Gelfand JR, Bookheimer SY. Dissociating neural mechanisms of temporal sequencing and processing phonemes. *Neuron* 2003;38:831–842. [PubMed: 12797966]
62. Buchsbaum BR, D'Esposito M. The search for the phonological store: from loop to convolution. *J Cogn Neurosci* 2008;20:762–778. [PubMed: 18201133]
63. Obleser J, Wise RJ, Alex Dresner M, Scott SK. Functional integration across brain regions improves speech perception under adverse listening conditions. *J Neurosci* 2007;27:2283–2289. [PubMed: 17329425]
64. Fu KMG, et al. Auditory cortical neurons respond to somatosensory stimulation. *J Neurosci* 2003;23:7510–7515. [PubMed: 12930789]
65. Kayser C, Petkov CI, Augath M, Logothetis NK. Functional imaging reveals visual modulation of specific fields in auditory cortex. *J Neurosci* 2007;27:1824–1835. [PubMed: 17314280]
66. Andersen RA, Buneo CA. Intentional maps in posterior parietal cortex. *Annu Rev Neurosci* 2002;25:189–220. [PubMed: 12052908]
67. Colby CL, Goldberg ME. Space and attention in parietal cortex. *Annu Rev Neurosci* 1999;22:319–349. [PubMed: 10202542]
68. Wise RJ, et al. Separate neural subsystems within 'Wernicke's area'. *Brain* 2001;124:83–95. [PubMed: 11133789]
69. 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. [PubMed: 12965041]
70. Warren JE, Wise RJ, Warren JD. Sounds do-able: auditory-motor transformations and the posterior temporal plane. *Trends Neurosci* 2005;28:636–643. [PubMed: 16216346]
71. Gibson, JJ. The theory of affordances. In: Shaw, R.; Bransford, J., editors. *Perceiving, Acting, and Knowing: Toward an Ecological Psychology*. Erlbaum; Hillsdale, New Jersey, USA: 1977. p. 67–82.
72. Rizzolatti G, Ferrari PF, Rozzi S, Fogassi L. The inferior parietal lobule: where action becomes perception. *Novartis Found Symp* 2006;270:129–140. discussion 140–125, 164–129. [PubMed: 16649712]
73. Hickok G, Poeppel D. Towards a functional neuroanatomy of speech perception. *Trends Cogn Sci* 2000;4:131–138. [PubMed: 10740277]
74. Hickok G, Poeppel D. The cortical organization of speech processing. *Nat Rev Neurosci* 2007;8:393–402. [PubMed: 17431404]
75. Jacquemot C, Scott SK. What is the relationship between phonological short-term memory and speech processing? *Trends Cogn Sci* 2006;10:480–486. [PubMed: 16997610]
76. Wilson SM, Saygin AP, Sereno MI, Iacoboni M. Listening to speech activates motor areas involved in speech production. *Nat Neurosci* 2004;7:701–702. [PubMed: 15184903]
77. Chen JL, Penhune VB, Zatorre RJ. Listening to musical rhythms recruits motor regions of the brain. *Cereb Cortex* 2008;18:2844–2854. [PubMed: 18388350]
78. Leaver AM, Van Lare JE, Zielinski BA, Halpern A, Rauschecker JP. Brain activation during anticipation of sound sequences. *J Neurosci* 2009;29:2477–2485. [PubMed: 19244522]
79. Müller-Preuss P, Ploog D. Inhibition of auditory cortical neurons during phonation. *Brain Res* 1981;215:61–76. [PubMed: 7260601]
80. Eliades SJ, Wang X. Neural substrates of vocalization feedback monitoring in primate auditory cortex. *Nature* 2008;453:1102–1106. [PubMed: 18454135]
81. Numminen J, Salmelin R, Hari R. Subject's own speech reduces reactivity of the human auditory cortex. *Neurosci Lett* 1999;265:119–122. [PubMed: 10327183]
82. Houde JF, Nagarajan SS, Sekihara K, Merzenich MM. Modulation of the auditory cortex during speech: an MEG study. *J Cogn Neurosci* 2002;14:1125–1138. [PubMed: 12495520]
83. Blakemore SJ, Wolpert DM, Frith CD. Central cancellation of self-produced tickle sensation. *Nat Neurosci* 1998;1:635–640. [PubMed: 10196573]
84. Guenther FH. Cortical interactions underlying the production of speech sounds. *J Commun Disord* 2006;39:350–365. [PubMed: 16887139]
85. Tourville JA, Reilly KJ, Guenther FH. Neural mechanisms underlying auditory feedback control of speech. *Neuroimage* 2008;39:1429–1443. [PubMed: 18035557]

86. Frey S, Campbell JS, Pike GB, Petrides M. Dissociating the human language pathways with high angular resolution diffusion fiber tractography. *J Neurosci* 2008;28:11435–11444. [PubMed: 18987180]
87. Petrides M, Pandya DN. Projections to the frontal cortex from the posterior parietal region in the rhesus monkey. *J Comp Neurol* 1984;228:105–116. [PubMed: 6480903]
88. Wolpert DM, Doya K, Kawato M. A unifying computational framework for motor control and social interaction. *Phil Trans R Soc Lond B* 2003;358:593–602. [PubMed: 12689384]
89. Grush R. The emulation theory of representation: motor control, imagery, and perception. *Behav Brain Sci* 2004;27:377–396. discussion 396–442. [PubMed: 15736871]
90. von Holst E, Mittelstaedt H. Das Reafferenzprinzip (Wechselwirkungen zwischen Zentralnervensystem und Peripherie). *Naturwissenschaften* 1950;37:464–476.
91. Hershberger W. Efference copy, the closed-loop analogue of von Holst's efference copy. *Cybernet Forum* 1976;8:97–102.
92. Jääskeläinen IP, et al. Human posterior auditory cortex gates novel sounds to consciousness. *Proc Natl Acad Sci USA* 2004;101:6809–6814. [PubMed: 15096618]
93. Bar M, et al. Top-down facilitation of visual recognition. *Proc Natl Acad Sci USA* 2006;103:449–454. [PubMed: 16407167]
94. Indefrey P, Levelt WJM. The spatial and temporal signatures of word production components. *Cognition* 2004;92:101–144. [PubMed: 15037128]
95. Seyfarth RM, Cheney DL, Marler P. Monkey responses to three different alarm calls: evidence of predator classification and semantic communication. *Science* 1980;210:801–803. [PubMed: 7433999]
96. Nishitani N, Hari R. Temporal dynamics of cortical representation for action. *Proc Natl Acad Sci USA* 2000;97:913–918. [PubMed: 10639179]
97. Suga N, O'Neill WE, Manabe T. Harmonic-sensitive neurons in the auditory cortex of the mustache bat. *Science* 1979;203:270–274. [PubMed: 760193]
98. Margoliash D, Fortune ES. Temporal and harmonic combination-sensitive neurons in the zebra finch's HVC. *J Neurosci* 1992;12:4309–4326. [PubMed: 1432096]
99. Rauschecker JP. Parallel processing in the auditory cortex of primates. *Audiol Neurootol* 1998;3:86–103. [PubMed: 9575379]
100. Burton H, Jones EG. The posterior thalamic region and its cortical projection in New World and Old World monkeys. *J Comp Neurol* 1976;168:249–301. [PubMed: 821975]

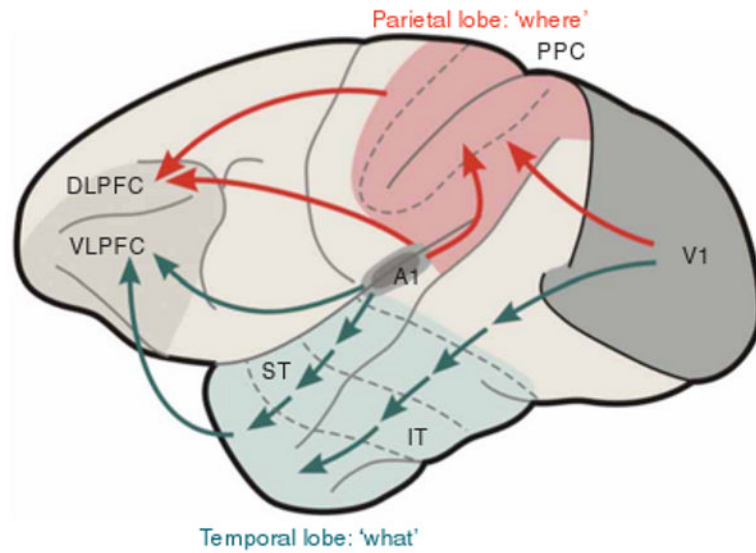


Figure 1. Dual processing scheme for 'what' and 'where', proposed for nonhuman primates on anatomical and physiological grounds. V1, primary visual cortex; A1, primary auditory cortex; IT, inferior temporal region; ST, superior temporal region; PPC, posterior parietal cortex; VLPFC, ventrolateral prefrontal cortex; DLPFC, dorsolateral prefrontal cortex. (Simplified from ^{refs. 4,5} and combined with an existing scheme from the visual system from ^{ref. 6}.)

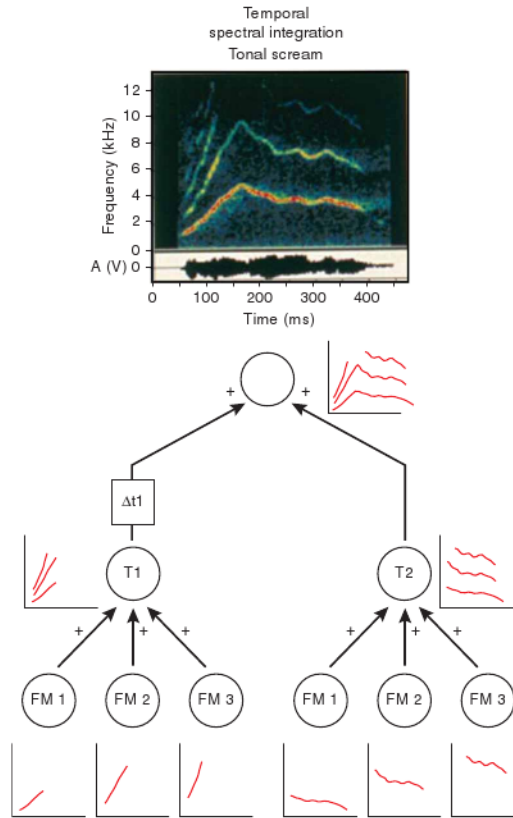


Figure 2.

Communication calls consist of elementary features, such as bandpass noise bursts or frequency-modulated (FM) sweeps. Harmonic calls, such as the vocal scream from the rhesus monkey repertoire depicted here by its spectrogram and time signal amplitude (A, measured as output voltage of a sound meter), consist of fundamental frequencies and higher harmonics. The neural circuitry for processing such calls is thought to consist of small hierarchical networks. At the lowest level, there are neurons serving as FM detectors tuned to the rate and direction of FM sweeps¹⁹; these detectors extract each FM component (shown in cartoon spectrograms) in the upward and downward sweeps of the scream. The output of these FM detectors is combined nonlinearly at the next level: the target neurons T1 and T2 possess a high threshold and fire only if all inputs are activated. At the final level, a ‘tonal-scream detector’ is created by again combining output from neurons T1 and T2 nonlinearly. Temporal integration is accomplished by having the output of T1 pass through a delay line with a latency Δt_1 sufficient to hold up the input to the top neuron long enough that all inputs arrive at the same time. Early processing of human speech sounds in the antero-lateral auditory belt and parabelt cortex is thought to be accomplished in a similar way⁹⁹.

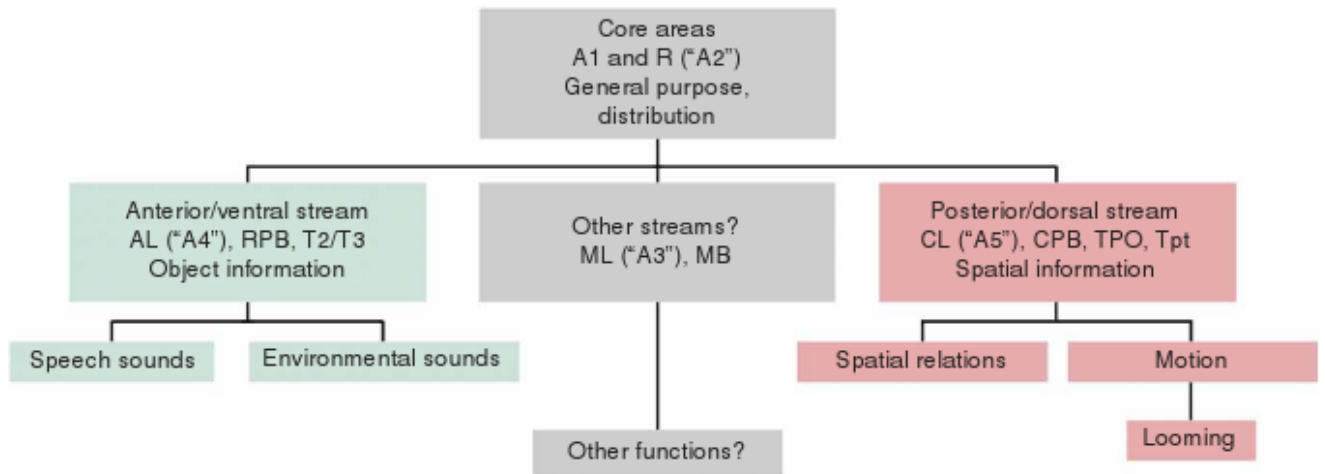


Figure 3.

Multiple parallel input modules advocated by some as an alternative to the dual-stream model. According to this model, sensory information at the cortical level originates from primary-like areas (A1 and R in the auditory system; R is also referred to as A2 by analogy to visual area V2) and splits into multiple early processing streams: an object stream (green) originating from the antero-lateral belt (AL; or “A4” by analogy to area V4, involved in processing visual form); a spatial stream (red) originating from the caudo-lateral belt (CL; or “A5” by analogy to visual motion area V5); and other streams or streamlets originating from either area ML between AL and CL (“A3” by analogy to visual area V3) or from the medial belt (MB). RPB and CPB, rostral and caudal parabelt; T2 and T3, temporal cortical areas as defined by Burton and Jones¹⁰⁰; TPO, polymodal cortex in the upper bank of superior temporal sulcus; Tpt, parieto-temporal area.

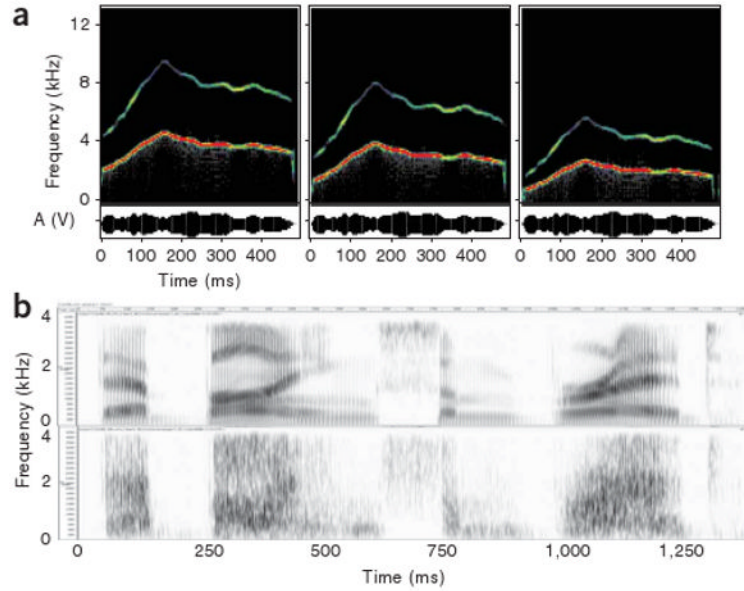
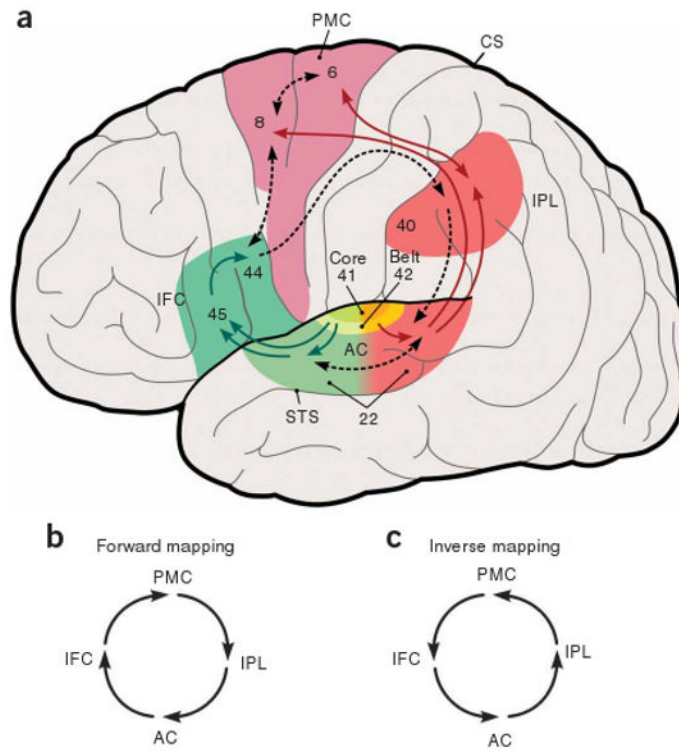


Figure 4.

Invariance in the perception of auditory objects (including vocalizations and speech) against transpositions in frequency, time or both. **(a)** Frequency-shifted monkey calls are behaviorally classified as the same by monkeys⁹⁵, presumably reflecting the response of higher-order neurons in anterior superior temporal cortex, even though the frequency contents of the monkey calls are markedly different. The example shows spectrograms of a tonal scream from a rhesus monkey frequency-shifted in steps of one octave. **(b)** Spectrograms of clear human speech (top) and of a six-channel noise-vocoded transformation of it⁴⁸ (bottom). The noise-vocoded version of the sentence (“They’re buying some bread.”) is easily comprehensible after short training, even though the sound is very impoverished in the spectral domain.

**Figure 5.**

Dual auditory processing scheme of the human brain and the role of internal models in sensory systems. This expanded scheme closes the loop between speech perception and production and proposes a common computational structure for space processing and speech control in the postero-dorsal auditory stream. **(a)** Antero-ventral (green) and postero-dorsal (red) streams originating from the auditory belt. The postero-dorsal stream interfaces with premotor areas and pivots around inferior parietal cortex, where a quick sketch of sensory event information is compared with a predictive efference copy⁹⁰ of motor plans. **(b)** In one direction, the model performs a forward mapping: object information, such as speech, is decoded in the antero-ventral stream all the way to category-invariant inferior frontal cortex (area 45), and is transformed into motor-articulatory representations (area 44 and ventral PMC), whose activation is transmitted to the IPL (and posterior superior temporal cortex) as an efference copy. **(c)** In reverse direction, the model performs an inverse mapping, whereby attention- or intention-related changes in the IPL^{66,67} influence the selection of context-dependent action programs in PFC and PMC. Both types of dynamic model are testable using techniques with high temporal precision (for example, magnetoencephalography in humans⁹⁶ or single-unit studies in monkeys) that allow determination of the order of events in the respective neural systems. AC, auditory cortex; STS, superior temporal sulcus; IFC, inferior frontal cortex, PMC, premotor cortex; IPL, inferior parietal lobule; CS, central sulcus. Numbers correspond to Brodmann areas.