



# Ventral and dorsal streams in the evolution of speech and language

Josef P. Rauschecker<sup>1,2\*</sup>

<sup>1</sup> Laboratory of Integrative Neuroscience and Cognition, Department of Neuroscience, Georgetown University Medical Center, Washington, DC, USA

<sup>2</sup> Mind/Brain Laboratory, Department of Biomedical Engineering and Computational Science, Aalto University, Helsinki, Finland

## Edited by:

Angela D. Friederici, Max Planck Institute for Human Cognitive and Brain Sciences, Germany

## Reviewed by:

Angela D. Friederici, Max Planck Institute for Human Cognitive and Brain Sciences, Germany  
Richard J. S. Wise, Imperial College London, UK

## \*Correspondence:

Josef P. Rauschecker, Laboratory of Integrative Neuroscience and Cognition, Department of Neuroscience, Georgetown University Medical Center, 3970 Reservoir Road, N.W., Washington, DC 20057-1460, USA.  
e-mail: rauschej@georgetown.edu

The brains of humans and old-world monkeys show a great deal of anatomical similarity. The auditory cortical system, for instance, is organized into a ventral and a dorsal pathway in both species. A fundamental question with regard to the evolution of speech and language (as well as music) is whether human and monkey brains show principal differences in their organization (e.g., new pathways appearing as a result of a single mutation), or whether species differences are of a more subtle, quantitative nature. There is little doubt about a similar role of the ventral auditory pathway in both humans and monkeys in the decoding of spectrally complex sounds, which some authors have referred to as auditory object recognition. This includes the decoding of speech sounds ("speech perception") and their ultimate linking to meaning in humans. The originally presumed role of the auditory dorsal pathway in spatial processing, by analogy to the visual dorsal pathway, has recently been conceptualized into a more general role in sensorimotor integration and control. Specifically for speech, the dorsal processing stream plays a role in speech production as well as categorization of phonemes during on-line processing of speech.

**Keywords:** cerebral cortex, macaque monkey, human, communication sounds, speech, music, internal models, brain connectivity

From an auditory point of view, spoken language starts with the processing of complex auditory signals. Physiological recordings in non-human primates suggest that neurons already at the secondary stage of processing along the auditory cortical pathway (the lateral belt areas) can show a preference for species-specific communication calls (Rauschecker et al., 1995). This response tuning is generated by convergence of input from lower-order neurons that respond to simple sounds like tones, frequency-modulated sweeps, or band-passed noise bursts. Neurons are sensitive to highly specific combinations of such inputs, and combining signals in a non-linear conjunctive AND-logic leads to the existence of neurons that respond specifically to certain types of calls. There is no reason to believe that the human auditory cortex does not contain similar neurons with combination sensitivity and a similar hierarchy from rather simple to more complex neurons, whose incidence increases from primary auditory cortex to more anterior regions of the superior temporal lobe (Rauschecker, 1998; Rauschecker and Tian, 2000).

Indeed, early studies of human auditory cortex with functional magnetic resonance imaging (fMRI) have shown that primary auditory cortex responds best to tones, while at the next stage, the equivalent of the lateral belt in the monkey, band-passed noise bursts are more effective stimuli (Wessinger et al., 2001). Further along the antero-ventral pathway, cortical regions are selectively activated by words and intelligible speech sounds (Binder et al., 2000; Scott et al., 2000). This hierarchical organization of the auditory ventral stream with regard to speech-sound processing was recently corroborated with more refined techniques

(Chevillet et al., 2011b). Furthermore, a meta-analysis of more than 100 neuroimaging studies of human speech processing has demonstrated that cortical regions in the mid-STG near the human lateral belt are sensitive to phonemes; farther afield in anterior STG, words are processed; finally, in the most anterior locations of STS, short phrases lead to selective activation (DeWitt and Rauschecker, 2012).

Invariant representation of sounds is another important step toward establishing a usable system for auditory communication, such as speech. There is evidence that invariances are formed along the antero-ventral stream as well (DeWitt and Rauschecker, 2012). However, other reports have found that premotor regions may be involved too (e.g., Chevillet et al., 2011a; Lee et al., 2012). It appears possible, therefore, that invariances are formed in different ways: once on the basis of spectro-temporal information, which is pooled along the frequency domain in the sense of an OR-logic within the auditory ventral stream; and independently in the domain of motor gestures, which are formed originally for speech production, but are invoked during the processing of speech as well. The same is almost inevitably true for the processing of other complex sounds that can be classified into discrete categories (Leaver and Rauschecker, 2010). Such auditory objects are also represented in anterior regions of the STG, but premotor cortex participates in their encoding as long as they can be produced and thus invoke a motor code. Monkeys are naturally handicapped by their less sophisticated vocal apparatus, which limits their vocal repertoire and their capacity to mimic sounds. The involvement of the dorsal pathway (including premotor regions)

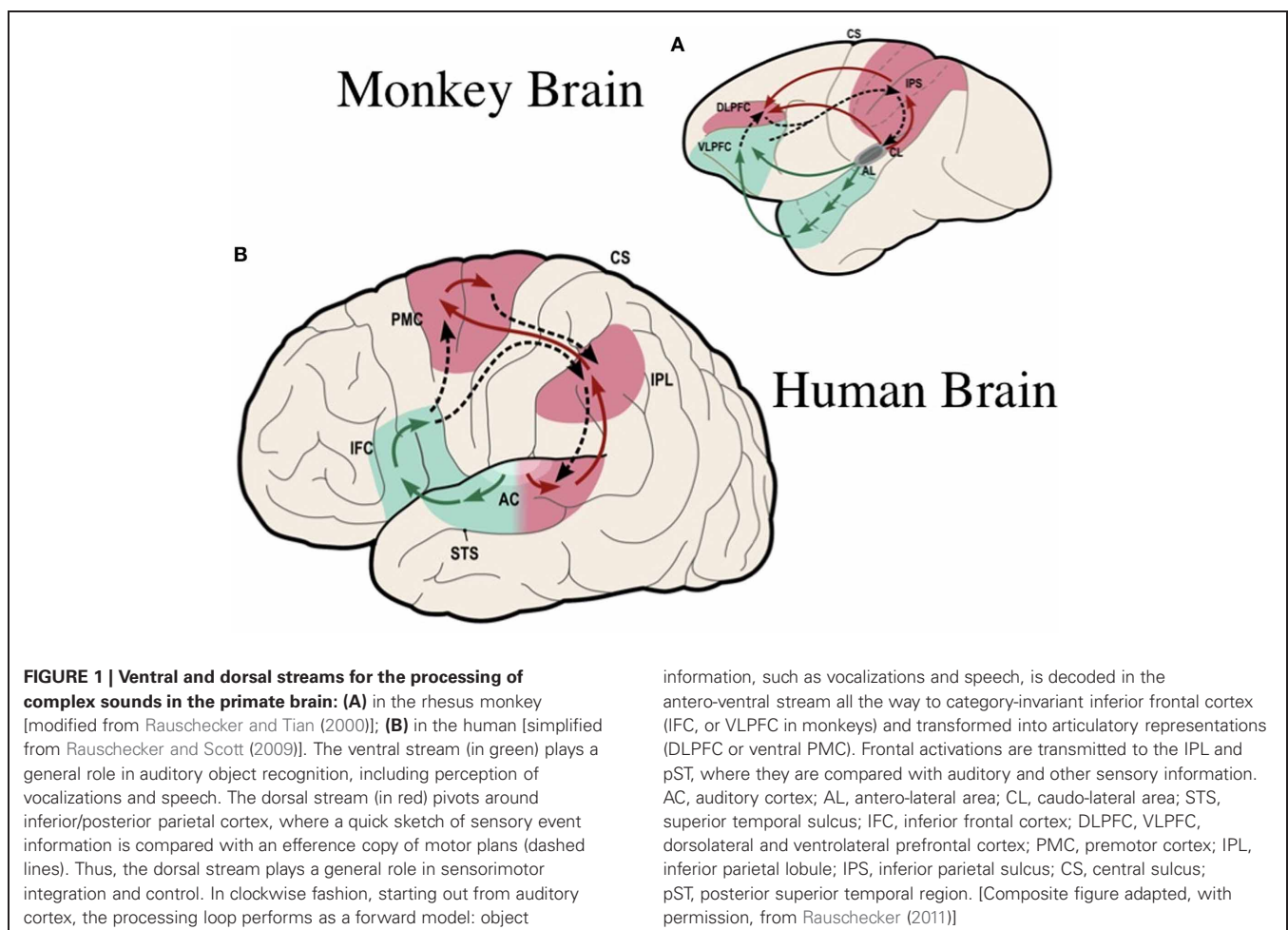
in the processing and categorization of self-produced sounds will, therefore, have to be tested by other means (Remedios et al., 2009).

The involvement of the dorsal auditory pathway, including premotor and inferior parietal regions, in the encoding and representation of temporally extended sounds (or sound sequences) became especially evident, when imagery of musical melodies was investigated (Leaver et al., 2009). During the learning of such sequences, the basal ganglia were actively engaged, whereas after these sequences became highly familiar, the same sequences activated more and more prefrontal areas. It appears, therefore, that the basal ganglia are responsible for the concatenation of sequential auditory information or formation of “chunks,” which represent information about conditional probabilities for one sound being followed by another. Once the chunks have been formed, they are once again stored in prefrontal regions. A similar chunking process occurs with cued sequences of learned finger movements (Koechlin and Jubault, 2006). This process involves prefrontal cortex near Broca’s area and has, therefore, been compared with models of language (Hagoort, 2005), redefining Broca’s area in terms of chunking (“unification”) of semantic, syntactic, and phonological information.

Thus, the role of the dorsal stream can be conceptualized into one of sensorimotor integration and control and applies to all

kinds of sequential stimuli, even beyond the auditory domain. Specifically for speech, the dorsal processing stream plays a role in speech production as well as categorization of phonemes during on-line processing of speech (Rauschecker and Scott, 2009; Rauschecker, 2011; **Figure 1**). The former role conforms to the classical idea of an “efference copy” or feed-forward model and allows for fast and efficient on-line control of speech production. By contrast, the latter function can be formalized as an inverse model during real-time speech processing, creating the affordances of the speech signal in a Gibsonian sense (Gibson, 1966; Rauschecker, 2005). Both functions require a (direct or indirect) connection between sensory and motor cortical structures of the brain, whereby subcortical structures (e.g., the basal ganglia) provide an additional link setting up transitional probabilities during associative learning of sound sequences.

Comparing human and monkey brain connectivity along the dorsal stream, there may be quantitative differences in the strengths of these connections, but there does not seem to be a difference in principle (Frey et al., 2008). Similarly, in the ventral stream, the fine-grain organization of cortical areas and the fine-tuning of its neuronal elements may be richer in humans than in monkeys, providing humans with a perceptual network for the detection of more subtle differences in the acoustic signal. The decisive distinction between humans and monkeys may,



however, lie in a third component where ventral and dorsal streams converge and interact: the prefrontal network. With its own hierarchical organization it provides the substrate for recursive processing of nested sequences, as they are typical for human grammatical language structures (Friederici, 2004). Again, however, this emergent new ability of humans may be based on a quantitative rather than principal difference in human and monkey brain organization, which ties in the existing strengths of both ventral and dorsal processing streams with fronto-parietal networks underlying working memory.

To test the real evolutionary similarity of human and monkey ventral and dorsal streams, two things have to happen in future studies:

(1) Connectivity studies in both species have to investigate in great detail which areas are connected. This will establish a greater amount of homology than other approaches, especially when the same techniques of structural and functional imaging are utilized. While anatomical tracer studies in monkeys will remain the gold standard (Romanski et al., 1999; Petrides and Pandya, 2009; Hackett, 2011), non-invasive fiber tractography using MRI-based technology will gain increasing importance as its resolution improves, because the exact same approach can be used in both species. Early attempts using diffusion tensor imaging (DTI) have had insufficient power to resolve crossing fibers within a single voxel or disentangle fibers with crossing trajectories (Catani et al., 2005; Croxson et al., 2005; Anwander et al., 2007; Rilling et al., 2008). Such studies have, therefore, remained inconclusive with regard to monkey-human homologies in language evolution. High-angular-resolution techniques, such

as diffusion spectrum imaging (DSI), have been utilized successfully in humans (e.g., Frey et al., 2008) and in monkeys (Schmahmann et al., 2007; Wedeen et al., 2008). Cross-validation studies of autoradiographic tract tracing and DSI in monkeys have shown a remarkable concordance of results between tracer studies and DSI (Schmahmann et al., 2007). However further improvements in resolution and reductions in scan time are certainly needed and possible, before DSI studies can become routine. Functional studies based on blood-oxygenation-level-dependent (BOLD) responses are feasible in both species as well (Petkov et al., 2006) and can elucidate connectivity to a certain extent. Microstimulation techniques as another approach to analyze connectivity (Kikuchi et al., 2008), on the other hand, are limited to animal studies.

(2) Behavioral monkey studies have to be designed that test the above concepts and go beyond traditional models. “What” and “where” processing are still characteristic for the two streams, but as generalized models are developed (Rauschecker and Scott, 2009; Rauschecker, 2011), more appropriate monkey studies have to follow. These studies have to focus on the computational transformations that occur between the various processing stages rather than merely the connectivity describing different anatomical pathways.

## ACKNOWLEDGMENTS

The collection of the underlying data as well as the writing of this article were supported by grants from the National Science Foundation (BCS-0519127, OISE-0730255), from the National Institutes of Health (R01NS052494, RC1DC010720), and from the Academy of Finland (FiDiPro).

## REFERENCES

- Anwander, A., Tittgemeyer, M., von Cramon, D. Y., Friederici, A. D., and Knösche, T. R. (2007). Connectivity-based parcellation of Broca's area. *Cereb. Cortex* 17, 816–825.
- Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S., Springer, J. A., Kaufman, J. N., and Possing, E. T. (2000). Human temporal lobe activation by speech and non-speech sounds. *Cereb. Cortex* 10, 512–528.
- Catani, M., Jones, D. K., and ffytche, D. H. (2005). Perisylvian language networks of the human brain. *Ann. Neurol.* 57, 8–16.
- Chevillet, M. A., Jiang, X., Rauschecker, J. P., and Riesenhuber, M. (2011a). Automatic phoneme categorization in the dorsal auditory pathway. *Soc. Neurosci. Abstr.* 172.09.
- Chevillet, M., Riesenhuber, M., and Rauschecker, J. P. (2011b). Functional localization of the ventral auditory “what” stream hierarchy. *J. Neurosci.* 31, 9345–9352.
- Croxson, P. L., Johansen-Berg, H., Behrens, T. E. J., Robson, M. D., Pinski, M. A., Gross, C. G., Richter, W., Richter, M. C., Kastner, S., and Rushworth, M. F. S. (2005). Quantitative investigation of connections of the prefrontal cortex in the human and macaque using probabilistic diffusion tractography. *J. Neurosci.* 25, 8854–8866.
- DeWitt, I., and Rauschecker, J. P. (2012). Phoneme and word recognition in the auditory ventral stream. *Proc. Natl. Acad. Sci. U.S.A.* 109, E505–E514.
- Frey, S., Campbell, J. S., Pike, G. B., and Petrides, M. (2008). Dissociating the human language pathways with high angular resolution diffusion fiber tractography. *J. Neurosci.* 28, 11435–11444.
- Friederici, A. D. (2004). Processing local transitions versus long-distance syntactic hierarchies. *Trends Cogn. Sci.* 8, 245–247.
- Gibson, J. J. (1966). *The Senses Considered as Perceptual Systems*. London: Allen and Unwin.
- Hackett, T. A. (2011). Information flow in the auditory cortical network. *Hear. Res.* 271, 133–146.
- Hagoort, P. (2005). On Broca, brain, and binding: a new framework. *Trends Cogn. Sci.* 9, 416–423.
- Kikuchi, Y., Rauschecker, J. P., Mishkin, M., Augath, M., Logothetis, N. K., and Petkov, C. I. (2008). Voice region connectivity in the monkey assessed with microstimulation and functional imaging. *Soc. Neurosci.* 34, 850.2.
- Koehlin, E., and Jubault, T. (2006). Broca's area and the hierarchical organization of human behavior. *Neuron* 50, 963–974.
- Leaver, A., van Lare, J. E., Zielinski, B. A., Halpern, A., and Rauschecker, J. P. (2009). Brain activation during anticipation of sound sequences. *J. Neurosci.* 29, 2477–2485.
- Leaver, A. M., and Rauschecker, J. P. (2010). Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. *J. Neurosci.* 30, 7604–7612.
- Lee, Y.-S., Turkeltaub, P., Granger, R., and Raizada, R. D. S. (2012). Categorical speech processing in Broca's area: an fMRI study using multivariate pattern-based analysis. *J. Neurosci.* 32, 3942–3948.
- Petkov, C. I., Kayser, C., Augath, M., and Logothetis, N. K. (2006). Functional imaging reveals numerous fields in the monkey auditory cortex. *PLoS Biol.* 4:e215. doi: 10.1371/journal.pbio.0040215
- Petrides, M., and Pandya, D. N. (2009). Distinct parietal and temporal pathways to the homologues of Broca's area in the monkey. *PLoS Biol.* 7:e170. doi: 10.1371/journal.pbio.1000170
- Rauschecker, J. P. (1998). Cortical processing of complex sounds. *Curr. Opin. Neurobiol.* 8, 516–521.
- Rauschecker, J. P. (2005). Vocal gestures and auditory objects. *Behav. Brain Sci.* 28, 143–144.
- Rauschecker, J. P. (2011). An expanded role for the dorsal auditory pathway in sensorimotor integration and control. *Hear. Res.* 271, 16–25.

- Rauschecker, J. P., and Scott, S. K. (2009). Maps and streams in the auditory cortex: non-human primates illuminate human speech processing. *Nat. Neurosci.* 12, 718–724.
- Rauschecker, J. P., and Tian, B. (2000). Mechanisms and streams for processing of “what” and “where” in auditory cortex. *Proc. Natl. Acad. Sci. U.S.A.* 97, 11800–11806.
- Rauschecker, J. P., Tian, B., and Hauser, M. (1995). Processing of complex sounds in the macaque nonprimary auditory cortex. *Science* 268, 111–114.
- Remedios, R., Logothetis, N. K., and Kayser, C. (2009). Monkey drumming reveals common networks for perceiving vocal and nonvocal communication sounds. *Proc. Natl. Acad. Sci. U.S.A.* 106, 18010–18015.
- Rilling, J. K., Glasser, M. F., Preuss, T. M., Ma, X., Zhao, T., Hu, X., and Behrens, T. E. J. (2008). The evolution of the arcuate fasciculus revealed with comparative DTI. *Nat. Neurosci.* 11, 426–428.
- Romanski, L. M., Tian, B., Fritz, J., Mishkin, M., Goldman-Rakic, P. S., and Rauschecker, J. P. (1999). Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nat. Neurosci.* 2, 1131–1136.
- Schmahmann, J. D., Pandya, D. N., Wang, R., Dai, G., D’Arceuil, H. E., de Crespigny, A. J., and Wedeen, V. J. (2007). Association fibre pathways of the brain: parallel observations from diffusion spectrum imaging and autoradiography. *Brain* 130, 630–653.
- Scott, S. K., Blank, C. C., Rosen, S., and Wise, R. J. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* 123, 2400–2406.
- Tian, B., Reser, D., Durham, A., Kustov, A., and Rauschecker, J. P. (2001). Functional specialization in rhesus monkey auditory cortex. *Science* 292, 290–293.
- Wedeen, V. J., Wang, R. P., Schmahmann, J. D., Benner, T., Tseng, W. Y., Dai, G., Pandya, D. N., Hagmann, P., D’Arceuil, H., and de Crespigny, A. J. (2008). Diffusion spectrum magnetic resonance imaging (DSI) tractography of crossing fibers. *Neuroimage* 41, 1267–1277.
- Wessinger, C. M., van Meter, J., Tian, B., van Lare, J., Pekar, J., and Rauschecker, J. P. (2001). Hierarchical organization of human auditory cortex revealed by functional magnetic resonance imaging. *J. Cogn. Neurosci.* 13, 1–7.
- was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 01 November 2011; paper pending published: 15 December 2011; accepted: 25 April 2012; published online: 15 May 2012.

Citation: Rauschecker JP (2012) Ventral and dorsal streams in the evolution of speech and language. *Front. Evol. Neurosci.* 4:7. doi: 10.3389/fnevo.2012.00007

Copyright © 2012 Rauschecker. This is an open-access article distributed under the terms of the Creative Commons Attribution Non Commercial License, which permits non-commercial use, distribution, and reproduction in other forums, provided the original authors and source are credited.

**Conflict of Interest Statement:** The author declares that the research