



OPEN ACCESS

EDITED BY

Marc Schönwiesner,
Leipzig University, Germany

REVIEWED BY

Sonja A. Kotz,
Maastricht University, Netherlands
Eliane Schochat,
University of São Paulo, Brazil
Erik Edwards,
Zeit Medical, Inc., United States

*CORRESPONDENCE

Sophie K. Scott
sophie.scott@ucl.ac.uk
Kyle Jasmin
kyle.jasmin@rhul.ac.uk

SPECIALTY SECTION

This article was submitted to
Auditory Cognitive Neuroscience,
a section of the journal
Frontiers in Neuroscience

RECEIVED 21 October 2022

ACCEPTED 28 November 2022

PUBLISHED 15 December 2022

CITATION

Scott SK and Jasmin K (2022)
Rostro-caudal networks for sound
processing in the primate brain.
Front. Neurosci. 16:1076374.
doi: 10.3389/fnins.2022.1076374

COPYRIGHT

© 2022 Scott and Jasmin. This is an
open-access article distributed under
the terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other
forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which
does not comply with these terms.

Rostro-caudal networks for sound processing in the primate brain

Sophie K. Scott^{1*} and Kyle Jasmin^{2*}

¹Institute of Cognitive Neuroscience, University College London, London, United Kingdom,

²Department of Psychology, Royal Holloway, University of London, Egham, United Kingdom

Sound is processed in primate brains along anatomically and functionally distinct streams: this pattern can be seen in both human and non-human primates. We have previously proposed a general auditory processing framework in which these different perceptual profiles are associated with different computational characteristics. In this paper we consider how recent work supports our framework.

KEYWORDS

speech perception, auditory cortex, neuroanatomy, auditory recognition, sensorimotor processing

“Hearing is a form of touch. You feel it through your body, and sometimes it almost hits your face”.

— Evelyn Glennie

“Intermittently she caught the gist of his sentences and supplied the rest from her subconscious, as one picks up the striking of a clock in the middle with only the rhythm of the first uncounted strokes lingering in the mind”.

— F. Scott Fitzgerald, *Tender is the Night*

Auditory processing in primates is neuroanatomically and functionally bifurcated. There are several models of speech and auditory processing in the human brain built around this principle (Alain et al., 2001; Hickok and Poeppel, 2004; Rauschecker and Scott, 2009; Jasmin et al., 2019), which originated in work on non-human primates (NHP). The NHP literature showed that rostral and caudal auditory cortical fields have distinctly different patterns of anatomical connectivity and different functional properties. For example, cells in rostral superior temporal sulcus were shown to be sensitive to the different kinds of non-human primate vocalizations (recognizing “monkey calls”) while those in the caudal fields were sensitive to the spatial location of the vocalizations (Rauschecker and Tian, 2000). These different functions have been described as “what” and “where/how” pathways within the rostral and caudal fields, respectively. Thus it was discovered that, in the visual system, auditory perception entails more than one kind of processing, with more than one functional goal.

This discovery was transformational for functional imaging studies of human speech processing, not just in terms of the neuroanatomical findings, but because it indicated

that different speech perception tasks might recruit different elements of the auditory perception network depending on the task. Tasks that required speech recognition networks, such as single word and sentence perception, consistently show recruitment of rostral temporal lobe fields (Mummery et al., 1999; Wise et al., 1999; Scott et al., 2000). By contrast, tasks that required motor engagement—e.g., speaking aloud, reading aloud in synchrony with other people (Jasmin et al., 2016), or when one's own voice is acoustically altered during speech production (Meekings and Scott, 2021), caudal auditory fields in humans are recruited. There is also a clear role for caudal auditory fields in representing the spatial location of voices (Hunter et al., 2002): all of these findings are consistent with a role for posterior auditory fields in guiding action.

Auditory neuroscience has made strides to move beyond mere description to computational mechanisms. Indeed, there have been significant advances in our understanding of the potential computational properties that underlie the functional differences seen in rostral/caudal auditory fields. In terms of anatomical connectivity, work by Scott et al. (2017) has shown convincingly that rostral and caudal auditory core, belt and parabelt areas receive different inputs from thalamic nuclei, which follows a caudal-rostral distinction: caudal auditory areas receive input mainly from the auditory thalamus and from the somatosensory thalamus (Hackett et al., 2007): moving rostrally, the medial geniculate body (the auditory thalamic input) drops, proportionally, and rostral auditory fields receive proportionally more input from the medial pulvinar, which receives input from the ascending visual pathway. Moving from caudal to rostral fields, the proportion of responses from subnuclei of the medial geniculate body also changes—from a ventral medial geniculate body dominance in caudal and mid-core auditory cortex, to a rough equivalence of inputs from the ventral medial geniculate body and the posterior dorsal medial geniculate body. Given the sheer complexity of the mammalian ascending auditory pathway, an important step in exploring the computational basis of different patterns of auditory processing is going to entail engaging with the nature of the representations of sound in these cortico-thalamic interactions. Some work on the stimulation of brain stem nuclei has suggested that there may even be processing pathways as early as the cochlear nucleus that have critical importance for speech perception (Moore and Shannon, 2009).

Scott et al. (2011) also showed that the caudal core field (A1) shows more detailed temporal response characteristics than the rostral temporal core area (RT): Neurons in caudal A1 respond faster to the onsets of sounds than rostral RT, and they are also accurate at tracking both fast and slow amplitude modulations. This stands in contrast to rostral RT, which responds more slowly to sound onsets and can only track slower amplitude modulations. Recent electrocorticography (ECoG) in humans are consistent with this macaque findings. Across human auditory cortex, regardless of the nature of the auditory

stimuli, the neural responses in caudal auditory fields are fast, transient, and linked to the onsets of sounds, while the neural responses in rostral auditory fields are slow and sustained (Hamilton et al., 2018). We argued in 2019 that these findings suggested a critical role for neuronal temporal responses in different kinds of computational processes on incoming sounds. In caudal fields, the responses to sound onsets are fast and temporally accurate, but not sustained, as responses that are critical to the control of action would need to be. By contrast, in rostral fields, the responses to sound onsets are slow and sustained, which potentially reflects hierarchical patterns of perceptual processing that interact with higher order linguistic and predictive processes.

This work has been recently replicated and extended in humans using fMRI. Zulfiqar et al. (2021) modeled fMRI BOLD responses for different temporal and spectral characteristics of the responses to stimuli. They found that caudal belt regions of the auditory cortex showed responses to natural sound stimuli that were fast but not frequency specific, responding to a broad spectral range. In contrast, rostral belt regions showed more specific spectral responses, and slower onset responses. Further support for this comes from another ECoG paper from Hamilton et al. (2021), which reported the shortest onset responses (generally less than 100 ms) in caudal Heschl's Gyrus (the location of primary auditory cortex in humans) and posterior superior temporal gyrus fields, and longer onset responses (up to 500 ms) in anterior superior temporal gyrus fields and the planum polare.

These findings strongly suggest that, as we hypothesized in 2019, the caudal/posterior “what/how” auditory pathway is underpinned by distinct computational processes from those of the anterior/rostral “what” pathway. Caudal fields (core and non-core) have responses that are generally fast, transient, and not necessarily specifically associated with particular stimulus characteristics: The responses in rostral fields (core and non-core) are generally slow and sustained and can be much more driven by stimulus specific properties. These distinctions are generalities—as can be seen in the Hamilton et al. (2021) paper, there is some overlap of these responses, but the general pattern is clear: Fast transient caudal responses reflect feed forward networks which are critical to the fast sensory guidance of action; slow, sustained responses in rostral fields likely reflect recognition processes which are slower as they require feedback processes from higher order language areas, which can have a profound effect on speech intelligibility (Obleser et al., 2007). This pattern reflects the overall cortical thickness gradient in the temporal lobes, such that primary auditory cortex is thin, with fewer feedback connections that cross cortical layers, whereas moving rostrally the cortex is thicker and has a higher ratio of feedback connections (Wagstyl et al., 2015).

Several studies have now shown that the rostral recognition “what” pathway, seen for intelligibility in speech, is not only seen for speech: music and other identifiable environmental sounds

also recruit the anterior temporal lobes in humans. There is compelling evidence that sound recognition is processed by parallel and distinct streams within these anterior fields (Norman-Haignere et al., 2015, 2022; Boebinger et al., 2021). This strongly suggests that while speech may often appear to dominate in these regions, that may be a function of the predominance of studies that focus on speech, and of the well-established speech processing problems that arise due to damage in left middle temporal artery territory. Using non-speech stimuli can show how speech fits within a wider range of auditory stimuli—in a recent ECoG study, song showed greater responses than speech or instrumental music within these fields (Norman-Haignere et al., 2022). However, a computational framework based on the temporal response properties we have described could be applied to a wide range of auditory stimuli—not necessarily specific to speech, as we have discussed (Jasmin et al., 2019). A challenge for further studies will be to determine the degree to which speech, song, instrumental music and other sound sources recruit distinct pathways, and what the computational properties are that may underlie these. This is all the more critical since there is good evidence that when we hear sounds in normal environments, they are rarely in silence, and rostral auditory areas seem to be key for simultaneously representing different sound sources (Evans et al., 2016).

These different auditory perceptual networks also interact with distributed systems throughout the human brain, including both other perceptual networks (including visual, somatosensory systems), and non-perceptual (including linguistic, emotional, musical networks): In many everyday auditory environments one would imagine that both auditory pathways are continually recruited. For example, during conversational speech, we have suggested that the rostral pathway is recruited to process the voice of the other speaker,

feeding into language networks that are also engaged in generating a response, while the caudal pathway is recruited to track the features of the other speaker's voice (e.g., the rate and the rhythm), such that the planned response is aligned with the talker's voice and a smooth turn taking can be managed (Scott et al., 2009). Auditory perception requires multiple kinds of perceptual processes, because the brain needs both to track the meaning of our auditory environments and to guide our production of sound into those environments.

Author contributions

Both authors listed have made a substantial, direct, and intellectual contribution to the work, and approved it for publication.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Alain, C., Arnott, S. R., Hevenor, S., Graham, S., and Grady, C. L. (2001). "What" and "where" in the human auditory system. *Proc. Natl. Acad. Sci. U.S.A.* 98, 12301–12306. doi: 10.1073/pnas.211209098
- Boebinger, D., Norman-Haignere, S., McDermott, J., and Kanwisher, N. (2021). Music-selective neural populations arise without musical training. *J. Neurophysiol.* 125, 2237–2263. doi: 10.1152/jn.00588.2020
- Evans, S., McGettigan, C., Agnew, Z., Rosen, S., and Scott, S. (2016). Getting the cocktail party started: Masking effects in speech perception. *J. Cogn. Neurosci.* 28, 483–500. doi: 10.1162/jocn_a_00913
- Hackett, T. A., De La Mothe, L. A., Ulbert, I., Karmos, G., Smiley, J., and Schroeder, C. E. (2007). Multisensory convergence in auditory cortex, II. Thalamocortical connections of the caudal superior temporal plane. *J. Comp. Neurol.* 502, 924–952. doi: 10.1002/cne.21326
- Hamilton, L. S., Edwards, E., and Chang, E. F. (2018). A spatial map of onset and sustained responses to speech in the human superior temporal gyrus. *Curr. Biol.* 28, 1860–1871. doi: 10.1016/j.cub.2018.04.033
- Hamilton, L. S., Oganian, Y., Hall, J., and Chang, E. F. (2021). Parallel and distributed encoding of speech across human auditory cortex. *Cell* 184, 4626–4639. doi: 10.1016/j.cell.2021.07.019
- Hickok, G., and Poeppel, D. (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition* 92, 67–99. doi: 10.1016/j.cognition.2003.10.011
- Hunter, M. D., Griffiths, T. D., Farrow, T. F., Zheng, Y., Wilkinson, I. D., Hegde, N., et al. (2002). A neural basis for the perception of voices in external auditory space. *Brain* 126, 161–169. doi: 10.1093/brain/awg015
- Jasmin, K. M., McGettigan, C., Agnew, Z. K., Lavan, N., Josephs, O., Cummins, F., et al. (2016). Cohesion and joint speech: Right hemisphere contributions to synchronized vocal production. *J. Neurosci.* 36, 4669–4680. doi: 10.1523/JNEUROSCI.4075-15.2016
- Jasmin, K., Lima, C. F., and Scott, S. K. (2019). Understanding rostral-caudal auditory cortex contributions to auditory perception. *Nat. Rev. Neurosci.* 20, 425–434. doi: 10.1038/s41583-019-0160-2
- Meekings, S., and Scott, S. K. (2021). Error in the superior temporal gyrus? A systematic review and activation likelihood estimation meta-analysis of speech production studies. *J. Cogn. Neurosci.* 33, 422–444. doi: 10.1162/jocn_a_01661
- Moore, D. R., and Shannon, R. V. (2009). Beyond cochlear implants: Awakening the deafened brain. *Nat. Neurosci.* 12, 686–691. doi: 10.1038/nn.2326

- Mummery, C. J., Ashburner, J., Scott, S. K., and Wise, R. J. (1999). Functional neuroimaging of speech perception in six normal and two aphasic subjects. *J. Acoust. Soc. Am.* 106, 449–457. doi: 10.1121/1.427068
- Norman-Haignere, S., Feather, J., Boebinger, D., Brunner, P., Ritaccio, A., McDermott, J., et al. (2022). A neural population selective for song in human auditory cortex. *Curr. Biol.* 32, 1470–1484. doi: 10.1016/j.cub.2022.01.069
- Norman-Haignere, S., Kanwisher, N., and McDermott, J. (2015). Distinct cortical pathways for music and speech revealed by hypothesis-free voxel decomposition. *Neuron* 88, 1281–1296. doi: 10.1016/j.neuron.2015.11.035
- Obleser, J., Wise, R. J., Dresner, M. A., and Scott, S. K. (2007). Functional integration across brain regions improves speech perception under adverse listening conditions. *J. Neurosci.* 27, 2283–2289. doi: 10.1523/JNEUROSCI.4663-06.2007
- Rauschecker, J. P., and Scott, S. K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nat. Neurosci.* 12, 718–724. doi: 10.1038/nn.2331
- 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. doi: 10.1073/pnas.97.22.11800
- Scott, B. H., Malone, B. J., and Semple, M. N. (2011). Transformation of temporal processing across auditory cortex of awake macaques. *J. Neurophysiol.* 105, 712–730. doi: 10.1152/jn.01120.2009
- Scott, B. H., Saleem, K. S., Kikuchi, Y., Fukushima, M., Mishkin, M., and Saunders, R. C. (2017). Thalamic connections of the core auditory cortex and rostral supratemporal plane in the macaque monkey. *J. Comp. Neurol.* 525, 3488–3513. doi: 10.1002/cne.24283
- 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. doi: 10.1093/brain/123.12.2400
- Scott, S., McGettigan, C., and Eisner, F. (2009). A little more conversation, a little less action—candidate roles for the motor cortex in speech perception. *Nat. Rev. Neurosci.* 10, 295–302. doi: 10.1038/nrn2603
- Wagstyl, K., Ronan, L., Goodyer, I. M., and Fletcher, P. C. (2015). Cortical thickness gradients in structural hierarchies. *Neuroimage* 111, 241–250. doi: 10.1016/j.neuroimage.2015.02.036
- Wise, R. J., Greene, J., Büchel, C., and Scott, S. K. (1999). Brain regions involved in articulation. *Lancet* 353, 1057–1061. doi: 10.1016/s0140-6736(98)07491-1
- Zulfiqar, I., Havlicek, M., Moerel, M., and Formisano, E. (2021). Predicting neuronal response properties from hemodynamic responses in the auditory cortex. *Neuroimage* 244:118575. doi: 10.1016/j.neuroimage.2021.118575