



Hierarchical processing for speech in human auditory cortex and beyond

Jonathan E. Peelle^{1*}, Ingrid S. Johnsrude² and Matthew H. Davis¹

¹ MRC Cognition and Brain Sciences Unit, Cambridge, UK

² Department of Psychology, Queen's University, Kingston, ON, Canada

*Correspondence: jonathan.peelle@mrc-cbu.cam.ac.uk

A commentary on

Hierarchical organization of human auditory cortex: evidence from acoustic invariance in the response to intelligible speech.

by Okada, K., Rong, F., Venezia, J., Matchin, W., Hsieh, I.-H., Saberi, K., Serences, J. T., and Hickok, G. (2010). *Cereb. Cortex*. doi:10.1093/cercor/bhp318.

The anatomical connectivity of the primate auditory system suggests that sound perception involves several hierarchical stages of analysis (Kaas et al., 1999), raising the question of how the processes required for human speech comprehension might map onto such a system. One intriguing possibility is that earlier areas of auditory cortex respond to acoustic differences in speech stimuli, but that later areas are insensitive to such features. Providing a consistent neural response to speech content despite variation in the acoustic signal is a critical feature of “higher level” speech processing regions because it indicates they respond to categorical speech information, such as phonemes and words, rather than idiosyncratic acoustic tokens. In a recent fMRI study, Okada et al. (2010) used multi-voxel pattern analysis (MVPA) to investigate neural responses to spoken sentences in canonical auditory cortex (i.e., superior temporal cortex), using a design modeled after Scott et al. (2000). Okada et al. (2010) used a factorial design that crossed speech clarity (clear speech vs. intelligible noise vocoded speech) with frequency order (normal vs. spectrally rotated). Noise vocoding reduces the amount of spectral detail in the speech signal but faithfully preserves temporal information. Depending on the reduction in spectral resolution (i.e., the number of bands used in vocoding), noise vocoded speech can be highly intelligible, especially following training. By contrast, spectral rotation of the speech signal renders it almost entirely unintelligible without any change

in overall level of spectral detail. Thus, the clear and vocoded sentences used by Okada et al. (2010) provided two physically dissimilar presentations of intelligible speech that the authors could use to identify acoustically insensitive neural responses; spectrally rotated stimuli allowed the authors to look for response changes due to intelligibility, independent of reductions in spectral detail.

In a standard whole-brain univariate analysis, Okada et al. (2010) found intelligibility-related responses (i.e., intelligible activity > unintelligible activity) in large portions of the superior temporal lobes bilaterally, as well as smaller activations in left inferior frontal gyrus, posterior fusiform gyrus, and premotor cortex. The authors then chose maxima for each participant within anatomically defined regions (bilateral posterior, middle, and anterior superior temporal sulcus [STS], as well as Heschl's gyrus) and performed MVPA analyses to assess the ability of these regions to discriminate among the four acoustic conditions. They found that Heschl's gyrus could reliably distinguish all conditions, despite showing a similar average hemodynamic response in the traditional mass univariate analysis. Regions of the STS showed varying degrees of sensitivity to acoustic information. In the left hemisphere, posterior STS was the most acoustically insensitive, followed by anterior STS and Heschl's gyrus (no reliable middle STS activation was identified). In the right hemisphere the greatest acoustic insensitivity was observed in middle STS, close to Heschl's gyrus, followed by anterior and posterior STS respectively. The authors interpret these findings as generally consistent with a hierarchical structure for speech processing in the temporal lobe, with regions of STS in both hemispheres playing a critical role in abstract phonological processes as indicated by their high acoustic insensitivity.

With these results, Okada et al. (2010) partially replicate previous univariate fMRI results reported by Davis and Johnsrude (2003). Davis and Johnsrude measured

neural activity in response to multiple speech conditions that were equally intelligible but differed acoustically. To achieve this, three different forms of speech degradation were employed: noise vocoded speech, speech segmented by noise bursts, and speech in continuous background noise. Within each type of speech degradation there were three matched levels of intelligibility (confirmed by pre-tests and behavioral ratings collected in the scanner). The authors first identified regions sensitive to speech intelligibility by correlating neural activity with behavioral performance, and then examined the degree to which each of these regions was also sensitive to the acoustic form of the stimuli. Activity in regions close to primary auditory cortex depended on the type of degradation, but other intelligibility-responsive regions were insensitive to this acoustic information. These acoustically insensitive areas included regions located anterior to peri-auditory areas bilaterally, posterior to left peri-auditory cortex, and left inferior frontal gyrus. This arrangement is broadly consistent with the anatomical organization of primate auditory cortex (Kaas et al., 1999) and suggests high levels of acoustic insensitivity in both anterior and posterior regions of left superior temporal cortex – consistent with the univariate analysis reported by Okada et al. (2010), but in contrast to their multivariate results. These conflicting findings regarding acoustic sensitivity in anterior temporal regions could be a result of either (a) the experimental design and specific stimuli used or (b) differing sensitivity of multivariate and univariate analysis methods, a question that requires further investigation.

Moving beyond the temporal lobe, the results of Davis and Johnsrude (2003) highlight the role of left inferior frontal cortex in speech comprehension. Activity in left inferior frontal cortex is common, although not universal, in neuroimaging studies of connected speech (e.g., Humphries et al., 2001; Davis and Johnsrude, 2003; Crinion and Price, 2005; Rodd et al., 2005, 2010; Obleser et al., 2007; Peelle et al., 2010b). Regions of

prefrontal cortex have extensive anatomical connections to auditory belt and parabelt regions (Hackett et al., 1999; Romanski et al., 1999) and are thus well positioned to modulate the operation of lower-level auditory areas. Davis and Johnsrude (2003) provided evidence linking this fronto-temporal modulation with the recovery of meaning from an impoverished acoustic signal by showing that inferior frontal responses were elevated for distorted-yet-intelligible speech compared to both clear speech and unintelligible noise. This result suggests that activation of inferior frontal regions is a neural correlate of the more effortful listening that is required for the comprehension of degraded speech.

Indeed, the relationship between intelligibility and listening effort also deserves consideration in interpreting temporal lobe responses to degraded speech. Okada et al. (2010) treated clear and vocoded speech as having similar intelligibility. This may be true in the sense that word report is equivalent (at ceiling); however, clear and vocoded conditions differ substantially in whether this intelligibility is achieved effortlessly (as for clear speech), or with considerable effort (for vocoded speech). In the Okada et al. (2010) study this difference in listening effort cannot be distinguished from sensitivity to acoustic

differences between stimuli. One way to control for these effects would be to match (below-ceiling) intelligibility across different types of acoustic degradation. Using this approach, Davis and Johnsrude (2003) observed that both inferior frontal and peri-auditory regions of the STS showed elevated signal for intelligible but degraded speech compared to both clear speech and noise. Like intelligibility-sensitive regions, the areas responding to listening effort demonstrated a hierarchical organization (i.e., differential degrees of acoustic sensitivity), and hence it might be that these two effects are confounded in temporal lobe responses observed by Okada.

On a related topic, we also note that Okada et al. (2010) used continuous fMRI, meaning that the auditory stimuli were presented in the midst of considerable background noise. Although one might assume that any such confounds would apply equally to all conditions tested, in fact, vocoded and clear sentences are not equally intelligible in the presence of background noise, even if word report scores are equivalent (at ceiling) when tested in quiet (Faulkner et al., 2001). Furthermore, even if participants are able to hear the sentences, scanner noise introduces significant additional task components related to segregating the auditory stream of interest

and downstream effects of listening effort (McCoy et al., 2005; Wingfield et al., 2006). To date effects of scanner noise have been associated with changes in neural activity using univariate approaches (Seifritz et al., 2006; Gaab et al., 2007; Peelle et al., 2010a); the effect on multivariate results is unknown. Generally, however, the results of auditory fMRI studies employing a standard continuous scanning sequence must be viewed with caution given the additional perceptual processes required.

Despite the caveats discussed above, how might the results of Okada et al. (2010) inform our understanding of speech processing? The authors suggest that their findings are consistent with a hierarchy for intelligible speech processing that starts with Heschl's gyrus, followed by anterior and posterior-going streams that progressively increase in acoustic invariance. In particular, their finding of acoustic sensitivity in anterior temporal regions stands in direct opposition to the original Scott et al. (2000) study and several follow-ups (Narain et al., 2003; Scott et al., 2006), as well as Davis and Johnsrude (2003), which argue that anterior temporal responses are largely acoustically invariant. Because of their focus on canonical regions of auditory cortex, Okada et al. (2010) do not discuss

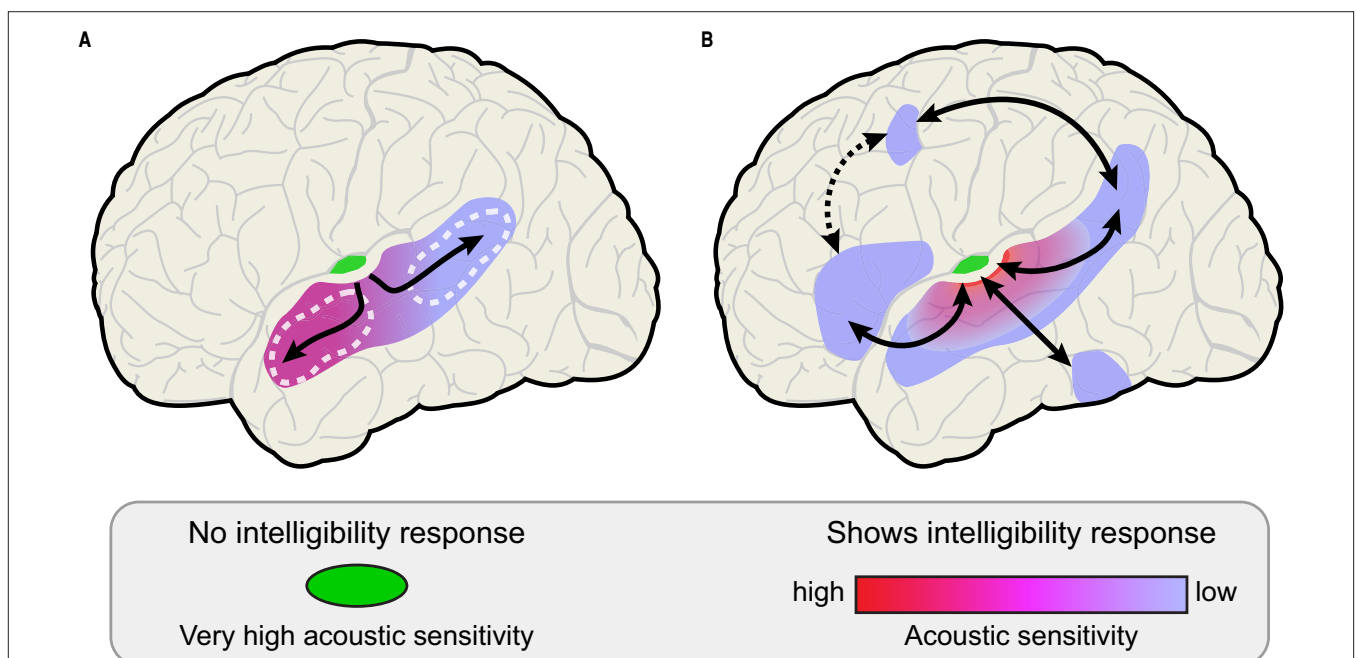


FIGURE 1 | Hierarchical models of processing intelligible speech. (A) Hierarchical processing in the temporal lobe, showing a posterior-anterior gradient in acoustic insensitivity moving away from primary auditory cortex.

Posterior and anterior regions of STS discussed by Okada et al. (2010) are outlined in white. **(B)** An expanded model of hierarchical processing for speech that includes prefrontal, premotor/motor, and posterior inferotemporal regions.

regions outside the superior temporal lobe as part of this hierarchy (Figure 1A).

Expanding on this description, we argue that a hierarchical model of speech comprehension necessarily includes regions of motor, premotor, and prefrontal cortex (Figure 1B) as part of multiple parallel processing pathways that radiate outward from primary auditory areas (Davis and Johnsrude, 2007). Electrophysiological studies in non-human primates demonstrate auditory responses in frontal cortex, suggesting not only strong frontal involvement, but that these regions may indeed be viewed as part of the auditory system (Kaas et al., 1999; Romanski and Goldman-Rakic, 2002). In addition to prefrontal cortex, left posterior inferotemporal cortex is also critically involved in the speech intelligibility network, especially in accessing or integrating semantic representations (Crinion et al., 2003; Rodd et al., 2005). Although anatomical studies in primates have long emphasized the extensive and highly parallel anatomical coupling between auditory and frontal cortices (Seltzer and Pandya, 1989; Hackett et al., 1999; Romanski et al., 1999; Petrides and Pandya, 2009), frontal regions have only recently become a prominent feature of models of speech processing (Hickok and Poeppel, 2007; Rauschecker and Scott, 2009). Unfortunately, discussions of auditory sentence processing often focus almost exclusively on the importance of superior temporal responses, even when frontal or inferotemporal activity is present (Humphries et al., 2001; Obleser et al., 2008; Okada et al., 2010), resulting in an incomplete picture of the neural mechanisms involved in speech comprehension.

In summary, there is now consensus that hierarchical processing is a key organizational aspect of the human cortical auditory system. The results of Okada et al. (2010) uniquely bring into question the degree to which anterior temporal cortex is acoustically insensitive, suggesting a more posterior locus for abstract phonological processing. Challenges for future studies include placing hierarchical organization in the temporal lobe within the broader context of larger networks for auditory and language processing, and clarifying the functional contribution of different parallel auditory processing pathways to comprehension of spoken language under varying degrees of effort.

ACKNOWLEDGMENTS

This work was supported by the Medical Research Council (U.1055.04.013.01.01). We are grateful to Greg Hickok for constructive comments on earlier drafts of this commentary.

REFERENCES

- Crinion, J., and Price, C. J. (2005). Right anterior superior temporal activation predicts auditory sentence comprehension following aphasic stroke. *Brain* 128, 2858–2871.
- Crinion, J. T., Lambon Ralph, M. A., Warburton, E. A., Howard, D., and Wise, R. J. S. (2003). Temporal lobe regions engaged during normal speech comprehension. *Brain* 126, 1193–1201.
- Davis, M. H., and Johnsrude, I. S. (2003). Hierarchical processing in spoken language comprehension. *J. Neurosci.* 23, 3423–3431.
- Davis, M. H., and Johnsrude, I. S. (2007). Hearing speech sounds: top-down influences on the interface between audition and speech perception. *Hear. Res.* 229, 132–147.
- Faulkner, A., Rosen, S., and Wilkinson, L. (2001). Effects of the number of channels and speech-to-noise ratio on rate of connected discourse tracking through a simulated cochlear implant speech processor. *Ear Hear.* 22, 431–438.
- Gaab, N., Gabrieli, J. D. E., and Glover, G. H. (2007). Assessing the influence of scanner background noise on auditory processing. II. An fMRI study comparing auditory processing in the absence and presence of recorded scanner noise using a sparse design. *Hum. Brain Mapp.* 28, 721–732.
- Hackett, T. A., Stepniewska, I., and Kaas, J. H. (1999). Prefrontal connections of the parabelt auditory cortex in macaque monkeys. *Brain Res.* 817, 45–58.
- Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8, 393–402.
- Humphries, C., Willard, K., Buchsbaum, B., and Hickok, G. (2001). Role of anterior temporal cortex in auditory sentence comprehension: an fMRI study. *Neuroreport* 12, 1749–1752.
- Kaas, J. H., Hackett, T. A., and Tramo, M. J. (1999). Auditory processing in primate cerebral cortex. *Curr. Opin. Neurobiol.* 9, 164–170.
- McCoy, S. L., Tun, P. A., Cox, L. C., Colangelo, M., Stewart, R., and Wingfield, A. (2005). Hearing loss and perceptual effort: downstream effects on older adults' memory for speech. *Q. J. Exp. Psychol.* 58, 22–33.
- Narain, C., Scott, S. K., Wise, R. J. S., Rosen, S., Leff, A., Iversen, S. D., and Matthews, P. M. (2003). Defining a left-lateralized response specific to intelligible speech using fMRI. *Cereb. Cortex* 13, 1362–1368.
- Obleser, J., Eisner, F., and Kotz, S. A. (2008). Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. *J. Neurosci.* 28, 8116–8124.
- Obleser, J., Wise, R. J. S., 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.
- Okada, K., Rong, F., Venezia, J., Matchin, W., Hsieh, I.-H., Saberi, K., Serences, J. T. and Hickok, G. (2010). Hierarchical organization of human auditory cortex: Evidence from acoustic invariance in the response to intelligible speech. *Cereb. Cortex*. doi: 10.1093/cercor/bhp318. [Epub ahead of print].

- Peelle, J. E., Eason, R. J., Schmitter, S., Schwarzbauer, C., and Davis, M. H. (2010a). Evaluating an acoustically quiet EPI sequence for use in fMRI studies of speech and auditory processing. *NeuroImage*. doi: 10.1016/j.neuroimage.2010.05.015. [Epub ahead of print].
- Peelle, J. E., Troiani, V., Wingfield, A., and Grossman, M. (2010b). Neural processing during older adults' comprehension of spoken sentences: Age differences in resource allocation and connectivity. *Cereb. Cortex* 20, 773–782.
- 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, e1000170. doi:10.1371/journal.pbio.1000170.
- 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.
- Rodd, J. M., Davis, M. H., and Johnsrude, I. S. (2005). The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. *Cereb. Cortex* 15, 1261–1269.
- Rodd, J. M., Longe, O. A., Randall, B., and Tyler, L. K. (2010). The functional organisation of the fronto-temporal language system: Evidence from syntactic and semantic ambiguity. *Neuropsychologia* 48, 1324–1335.
- Romanski, L. M., and Goldman-Rakic, P. S. (2002). An auditory domain in primate prefrontal cortex. *Nat. Neurosci.* 5, 15–16.
- 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.
- Scott, S. K., Blank, C. C., Rosen, S., and Wise, R. J. S. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* 123, 2400–2406.
- Scott, S. K., Rosen, S., Lang, H., and Wise, R. J. S. (2006). Neural correlates of intelligibility in speech investigated with noise vocoded speech—A positron emission tomography study. *J. Acoust. Soc. Am.* 120, 1075–1083.
- Seifritz, E., Di Salle, F., Esposito, F., Herdener, M., Neuhoff, J. G., and Scheffler, K. (2006). Enhancing BOLD response in the auditory system by neurophysiologically tuned fMRI sequence. *Neuroimage* 29, 1013–1022.
- Seltzer, B., and Pandya, D. N. (1989). Frontal lobe connections of the superior temporal sulcus in the rhesus monkey. *J. Comp. Neurol.* 281, 97–113.
- Wingfield, A., McCoy, S. L., Peelle, J. E., Tun, P. A., and Cox, C. (2006). Effects of adult aging and hearing loss on comprehension of rapid speech varying in syntactic complexity. *J. Am. Acad. Audiol.* 17, 487–497.

Received: 09 April 2010; accepted: 28 May 2010; published online: 28 June 2010.

Citation: Peelle JE, Johnsrude IS and Davis MH (2010) Hierarchical processing for speech in human auditory cortex and beyond. *Front. Hum. Neurosci.* 4:51. doi: 10.3389/fnhum.2010.00051

Copyright © 2010 Peelle, Johnsrude and Davis. This is an open-access article subject to an exclusive license agreement between the authors and the Frontiers Research Foundation, which permits unrestricted use, distribution, and reproduction in any medium, provided the original authors and source are credited.