From speech and talkers to the social world:
The neural processing of human spoken language

Sophie K. Scott

Human speech perception is a paradigm example of the complexity of human linguistic processing; however, it is also the dominant way of expressing vocal identity and is critically important for social interactions. Here, I review the ways that the speech, the talker, and the social nature of speech interact and how this may be computed in the human brain, using models and approaches from nonhuman primate studies. I explore the extent to which domain-general approaches may be able to account for some of these neural findings. Finally, I address the importance of extending these findings into a better understanding of the social use of speech in conversations.

Speech is often viewed as the auditory form of a more abstract linguistic system and from a neuropsychological perspective, the processing of this spoken language has been associated with the left posterior temporal lobe (e.g., 1, 2). However, this view has been challenged, first, by models of auditory processing from the nonhuman primate literature arguing for multiple anatomical and functional streams of processes in speech and sound perception (2); second, by studies of the processing of the nonlinguistic information encoded in the speaking voice (3); and third, by the suggestion that we need to bear in mind the critical social importance of speech (4). From the first perspective, the involvement of the dorsolateral temporal lobes in auditory speech processing may rest upon perceptual networks consisting of multiple processing pathways, comparable to the distinct processing streams seen in the visual system. From the second perspective, if there is speech, there is always a talking voice (5) and there may be significant hemispheric asymmetries involved in the ways that left and right temporal fields process these different kinds of information (6, 7). From the third perspective, we can study speech in isolation, but the neural systems that process speech develop these skills in social, conversational settings, and speech is an overwhelmingly social behavior.

Speech: From sound to meaning

The human speech signal is almost bewilderingly complex in terms of acoustics. Speech sounds are made by a variety of different actions, from brief, bursting releases to lengthy, noisy segments and from nasal sounds to sustained vowels (8). Although all speech depends on detailed spectrotemporal processing, the precise acoustic cues that have linguistic relevance vary somewhat from language to language. When confronted with a phoneme or an acoustic cue that is not used in our own language, we can struggle to hear this accurately and to use this linguistically (9). This problem comes about because of the development of phonetic processing skills in the acquisition of speech perception (10). The developing brain learns to prioritize the auditory cues that are linguistically relevant and to downgrade the importance of speech sounds that are less important for distinguishing the meaning of words. Indeed, the processing of words in a holistic fashion may come before more fine-grained phonological skills in development (11).

However, human speech perception is complex and multi-stable; in the brain of a listener, no one acoustic cue determines the intelligibility of speech, and listeners will listen flexibly, making use of relevant acoustic cues when they are of utility (12, 13). This flexibility is necessary because we are continually confronted with new voices and accents, as well as

complex acoustic listening environments. Auditory cortical fields therefore need to adapt in a transient way to both speech and the listening conditions in which speech is heard. Furthermore, although speech is made up of sequences of speech sounds, the speech sounds themselves are affected by the sounds around them. Speech sounds are different depending upon their position in a syllable—for most British- English accents, the "l" at the start of "leaf" is very different from the "l" at the end of "bill." Listeners are very sensitive to this contextual information and use it to help them decode speech (14), and the recognition of spoken words is based not on a sequence of abstract phonemic categories, but more on a form- based auditory representation (15, 16).

Speech is the auditory form of language, and much of the aim of research has been to study the deeper structures contained with the computational properties of language, with the surface structure of the speech being some- what dissociable from this higher-order in-formation. However, the acoustic form of speech is rich in terms of information, from phonemic, talker, emotion, and effects through to higher-order information such as grammatical structure.

Speech is also rarely heard in silence; we commonly listen to speech in noisy environments. Adults struggle most to understand speech in the context of broadband noisy sounds (such as an air-conditioning unit or a fan) (17). However, the more "information" there is in competing speech, the more likely it is to be processed by the adult listener, especially if it starts to compete with the attended speech in terms of its semantic content (17).

Although it is conceptually simple to separate speech production and speech perception mechanisms, perception networks are critical to the production of speech. Even moderate patterns of hearing loss in development will have an effect on the development of speech skills (18); by contrast, the absence of the ability to speak aloud is not detrimental to the ability to learn to understand speech (19). Speech production skills, therefore, like many controlled motor skills, are dependent on perceptual processing of one's own actions (20). One final complication is that human speech expresses the phenomenal complexity of hu-man language, but it is also a primarily social behavior: we very rarely speak aloud on our own, and human speech is universally the dominant mode for social engagement (21). Conversations are the contexts in which we learn to speak (22) and are the settings for most social interactions. The study of speech thus lies at the crux of auditory processing, language processing, social processing, emotion, identity, and music. How have developments in neuroscience helped us to model this immense complexity?

The dorsolateral temporal lobes

Speech is an auditory signal and it is thus processed in the ascending auditory pathway, to the primary auditory cortex (PAC) and surrounding auditory association cortex, and extending out laterally and down into the superior temporal gyrus (STG). In nonhuman primates, the PAC is formed of three core fields, arranged in a caudal-rostral orientation (23, 24), and these then project to surrounding belt and parabelt fields (25). These projections maintain the rostral-caudal structure of the core fields, and this connectivity is preserved in the projections to frontal fields. There are functional differences seen within these rostral-caudal fields in nonhuman primates, with rostral fields sensitive to different kinds of conspecific vocalizations and caudal fields sensitive to the spatial locations of the sounds (26) and to somatosensory stimulation (27). Thus, perceptual processing in primates is not one unitary phenomenon, but rather one that rests upon different perceptual networks, which can be recruited differentially depending on the task. In the real world, actions will depend on these networks working together.

Quite obviously, nonhuman primates do not speak; they can make use of complex vocalizations and may have similar laryngeal flexibility (28), but their vocalizations do not verge on the complexity of human speech and language. Is there any value, therefore, in trying to map from these nonhuman primate studies onto models of human speech processing?

First, the recruitment of rostral auditory fields in speech perception has been widely reported: unlike the left posterior superior temporal sulcus fields, which had been emphasized as the core region associated with Wernicke's aphasia, early functional imaging studies of speech perception revealed a dominant response to speech that ran forward from PAC toward the temporal pole, into the anterior superior temporal gyrus (29–33). These left anterior STS fields are sensitive to intelligible speech no matter what the talker sounds like (29, 31, 33), and rostral STG/STS fields show selective responses to phonetic, syntactic, and semantic information (34–37). There is a highly sensitive pattern of the processing of phonemic sequences with the STG, which displays the flexibility and adaptability that these networks require to navigate a world of talkers with different accents and different listening environments (38). This flexibility must also apply to the competing auditory environment; studies in which participants listen to speech while ignoring competing auditory maskers show responses within rostral temporal fields that are highly sensitive to the masking sounds (39–41). Unattended competing talkers' voices are not discarded at the auditory periphery, but rather are processed within the same network as the attended speech. This permits the processing of unattended speech for some potentially relevant information, but also means that it is competing for resources with the attended speech (17, 42). This must mean that rostral auditory fields are capable of the creation and representation of patterns of speech that are being heard at the same times as the attended speech. Given the complexity of auditory environments, this may simply be a glimpse of the nature of the formation and maintenance of these rostral auditory representations and our ability to switch our attention between them (43).

These recognition processes within rostral fields show some important hemispheric asymmetries. Right rostral temporal fields are sensitive to voice-specific information in non-human primates (44) and are also highly sensitive to natural pitch profiles (intonation) in speech (45, 46). Right rostral fields dominate in the processing of talker identity in humans (47), although this can also show bilateral responses. Humans rely strongly on pitch cues to discriminate between talkers (48). These differences between linguistic and nonlinguistic processing of vocal information in the left and right rostral fields do not rest on basic auditory processing differences (6, 7), but rather seem to reflect differences in the kind of information being processed.

Listeners use the accent of talkers to help then interpret word meanings; e.g., British-English listeners will be more likely to interpret "bonnet" as meaning "hat" if it is said in an American accent (49). Talkers who are familiar are easier for us to understand, and listeners adapt very quickly to talker-specific quirks of speech production, showing phoneme-specific adaptations, but only to that speaker (50, 51).

We are also much more accurate at discriminating talkers in a language that we speak; in a language we do not speak, we can struggle to tell talkers apart (52). These studies must mean that speech and talker recognition brain networks may be distinct anatomically, but must interact quickly, continuously, and accurately (Fig. 1). For example, noninvasively F1 altering function in left STG fields with transcranial direct current stimulation disrupts adaptation to the speech of talkers (53).

By contrast, caudal auditory fields are much less sensitive to the specific kind of speech and voice information being processed, and more sensitive to their sensorimotor associations. Caudal auditory regions are reliably activated when people move their articulators to make a sound, or even if they mimic these actions silently (54–56). This seems to map onto a key role for caudal auditory areas in the sensory guidance of speech and voice production. When people change their voice because they are coping with altered perceptual consequences of speaking, the detection of and compensation for these alterations are associated with increased recruitment of caudal auditory fields (57). Aligning your voice with another talker in unison speech (58) or deliberate controlled production of a different vocal identity when speaking (i.e., trying to sound like someone else) also recruits these caudal areas (59).

The nature of auditory representations of speech

The importance of phonemes in speech can lead to the assumption that they must form important representational cues in the perceptual processing of speech. However, this does not mean that phonemes themselves, as discrete, abstract linguistic individual items, are encoded during speech comprehension before any lexical processing; indeed, one might well expect that the sequential dependencies be- tween speech sounds, which are highly in- formative to listeners, would not be discarded in the processes that lead to the comprehension of speech. There is evidence for early perceptual processing of phonological information (60) and its syntactic influences (61) but that does not mean that individual phonemes are being represented as discrete, abstract items. STG is highly sensitive to the introduction of longer phonetic sequences, but less sensitive to the introduction of clusters of phonemes (62).

The nature of phonemes in the STG seems to be more important in how they contribute to the phone- mic "shape" of a sequence and that this form is represented at the syllable level, rather than at the level of individual phonemes. Phonemes contribute to the sequence like facial features contribute to face: their perceptual role contributes to the whole, rather than to the assembled individual items. Studies of the temporal sensitivity in STG fields show a relatively slow sensitivity, peaking at time scales associated with words or syllables, rather than the much faster time scales associated with individual phonemes (63) (Fig. 2). Similarly, electrocorticogram (eCOG) data show a sensitivity to the amplitude envelope of speech sounds, which correlates broadly with syllable structure (64).

Syllables are also good candidates for linguistic universals in the organization of spoken language (65), and syllable structure is quite constrained. At its bare minimum, a single syllable can consist of one vowel, and the ways that consonants can be added to the onset and offset of the vowel varies across languages, although the simplest structure around the world is consonant-vowel or CV, not VC. English permits highly complex syllables with up to three consonants before the vowel and four consonants after, e.g., "strengths," which has a $C_1C_2C_3VC_1C_2C_3C_4$ structure. In Japanese, such sequences of clusters of consonants would be illegal, and syllables have a much simpler CVC structure. It is clear that syllable structure gives a frame that can be roughly construed as corresponding to the onset or rhyme of the syllable, and which might form the basic (and language specific) context in which the phoneme sequences are represented. When presented with consonant clusters, Japanese listeners hear vowels between the consonants, called epenthetic vowels, and the perception of these epenthetic vowels is linked to experience-dependent processing in the STG (66). Perhaps part of the language- specific acoustic phonetic processing that is seen within the STG may form these syllable- level constraints upon speech perception.

Are we really looking at a speech-specific system?
To what extent will any of these networks be specific to speech and language? The distinction between linguistic and nonlinguistic aspects of sound processing may not be fully independent. The phonetic "tuning" into language-relevant aspects of sounds that happens as we acquire a spoken language during development may also affect the perceptual processing of non-speech sounds (9). In terms of neural activation, there may not be a complete functional dissociation of speech and non- speech processing.

However, it is almost certainly not the case that any of these networks are limited to the processing of vocal information. If one limits oneself to questions about speech and language, then one will only find results that relate to that question; however, this approach has not had great success in divining the computational principles that might distinguish the rostral and caudal streams. We may need to reframe the rostral–caudal auditory distinction as a more general one that reflects rostral recognition processes, not one limited to linguistic information (or, indeed, to vocal information) (67). In parallel, caudal auditory networks may not be solely involved in coordinating speech-related sensorimotor links. Rostral and caudal auditory fields may be distinguished on the basis of their temporal response profiles: rostral fields show a slow, sustained response to sounds, whereas caudal fields show a fast, transient response to the onsets of sounds (38, 68). The slower rostral pathway responses would be consistent with feedback influences on recognition processes, whereas the faster caudal responses would be consistent with the fast- online perceptual guidance of action (Fig. 3).

Consistent with the suggestion of a more domain-general approach to auditory processing, there are distinct responses to the different classes of auditory objects and these lie within rostral auditory fields (69). By contrast, caudal fields are recruited when participants listen to sounds that contain information about the locations of actions and objects in space and their proximity to the listener (70–72).

Speech in a social world
Social engagement with speech and language is seen in a further linguistic universal: spoken conversation. Speech is the dominant mode for social interactions in almost all human cultures, apart from those for whom signing is the dominant mode. As a field, we have prioritized the acoustic, linguistic, and computational aspects of speech processing over its social relevance. Accounting for at least some of the cortical activation seen in the perceptual processing of speech is that outside of the laboratory, we encounter speech in social interactions, interactions in which we are commonly active participants (4, 73). A specific role for premotor fields in the tracking of the rhythm of speech was suggested as a candidate way of aligning one's own speech with other talkers in a conversation and facilitating the swift changes of talker turns that characterize spoken conversations. A recent study (63) looked at the sensitivity to time scales in spoken language and found an important role for premotor cortex in the processing of phrasal information in spoken sequences. By contrast, STG fields were more sensitive to the structure of syllable range sequences (Fig. 2). This suggests that premotor cortex is sensitive to information at the kind of scale that could be important for rhythmic alignment in spoken language. Comparing face-to-face conversation with face- to-face repetition reveals greater whole-brain involvement during conversation in frontal and temporal areas, with an emphasis on the left temporal pole, left temporoparietal junction, and bilateral medial prefrontal cortex (74). This suggests that face-to-face conversation indeed recruits different elements of the

speech perception network, including both rostral and caudal networks, alongside other networks recruited in face processing.

It is going to be critical to understand the ways that the human brain is engaged in spoken communication and the social and emotional impact of this. Speech is more than language and it seems that this may be more important to the healthy function of humans and their brains than previously thought. The goals for further neuroscientific investigations of speech perception should start to address the ways that we engage with speech, from sound to social meaning and social contact.

REFERENCES AND NOTES

1. J. E. Bogen, G. M. Bogen, Ann. N. Y. Acad. Sci. 280, 834–843 (1976).
2. J. P. Rauschecker, S. K. Scott, Nat. Neurosci. 12, 718–724 (2009).
3. R. Watson, M. Latinus, I. Charest, F. Crabbe, P. Belin, Cortex 50, 125–136 (2014).
4. S. K. Scott, C. McGettigan, F. Eisner, Nat. Rev. Neurosci. 10, 295–302 (2009).
5. S. K. Scott, C. McGettigan, in APA Handbook of Nonverbal Communication, D. Matsumoto, H. Hwang, M. Frank, Eds. (American Psychological Association, 2015), pp. 289–306.
6. C. McGettigan, S. K. Scott, Trends Cogn. Sci. 16, 269–276 (2012).
7. S. K. Scott, C. McGettigan, Brain Lang. 127, 36–45 (2013). 8. P. Ladefoged, P. Maddieson, The Sounds of the World's Languages (Blackwell, 1996).
9. P. Iverson, A. Wagner, S. Rosen, J. Acoust. Soc. Am. 139, 1799–1809 (2016).
10. J. F. Werker et al., Cognition 103, 147–162 (2007).
11. M. M. Vihman, Br. J. Psychol. 108, 1–27 (2017).
12. G. A. Miller, Language and Communication (McGraw Hill, 1951).
13. R. V. Shannon, F. G. Zeng, V. Kamath, J. Wygonski, M. Ekelid, Science 270, 303–304 1995).
14. S. Hawkins, J. Phonetics 31, 373–405 (2003).
15. J. E. Andruski, S. E. Blumstein, M. Burton, Cognition 52, 163–187 (1994).
16. B. McMurray, M. K. Tanenhaus, R. N. Aslin, Cognition 86, B33–B42 (2002).
17. D. S. Brungart, B. D. Simpson, M. A. Ericson, K. R. Scott, J. Acoust. Soc. Am. 110, 2527–2538 (2001).
18. K. Mogford, "Oral language acquisition in the prelinguistically deaf" in Language Development in Exceptional Circumstances, D. V. M. Bishop and K. Mogford, Eds. (Churchill Livingstone, 1988), pp. 110–131.
19. D. V. M. Bishop, in Language Development in Exceptional Circumstances, D. V. M. Bishop and K. Mogford, Eds. (Churchill Livingstone, 1988), pp. 220–238.
20. A. A. Mattar, S. M. Nasir, M. Darainy, D. J. Ostry, Prog. Brain Res. 191, 31–44 (2011).
21. G. A. Miller, Review of Greenberg JL, Universals of Language, Contemporary Psychology 8, 417–418 (1963).
22. E. V. Clark, Lang. Learn. Dev. 14, 170–185 (2018).
23. J. H. Kaas, T. A. Hackett, Proc. Natl. Acad. Sci. U.S.A. 97, 11793–11799 (2000).
24. B. H. Scott et al., Cereb. Cortex 27, 809–840 (2017).
25. L. M. Romanski et al., Nat. Neurosci. 2, 1131–1136 (1999).
26. B. Tian, D. Reser, A. Durham, A. Kustov, J. P. Rauschecker, Science 292, 290–293 (2001).
27. T. A. Hackett et al., J. Comp. Neurol. 502, 924–952 (2007).
28. W. T. Fitch, B. de Boer, N. Mathur, A. A. Ghazanfar, Sci. Adv. 2, e1600723 (2016).
29. S. K. Scott, C. C. Blank, S. Rosen, R. J. S. Wise, Brain 123, 2400–2406 (2000).
30. C. McGettigan et al., J. Cogn. Neurosci. 24, 636–652 (2012).

31. K. Okada et al., Cereb. Cortex 20, 2486–2495 (2010).

32. S. Evans et al., Cereb. Cortex 24, 2350–2361 (2014).

33. S. Rosen, R. J. S. Wise, S. Chadha, E. J. Conway, S. K. Scott, PLOS ONE 6, e24672 (2011).

34. Z. K. Agnew, C. McGettigan, S. K. Scott, J. Cogn. Neurosci. 23, 4038–4047 (2011).

35. Z. K. Agnew, H. van de Koot, C. McGettigan, S. K. Scott, Lang. Cogn. Neurosci. 29, 1035–1045 (2014).

36. L. Cohen, A. Jobert, D. Le Bihan, S. Dehaene, Neuroimage 23, 1256–1270 (2004).

37. A. D. Friederici, S. A. Kotz, S. K. Scott, J. Obleser, Hum. Brain Mapp. 31, 448–457 (2010).

38. H. G. Yi, M. K. Leonard, E. F. Chang, Neuron 102, 1096–1110 (2019).

39. S. Evans, C. McGettigan, Z. K. Agnew, S. Rosen, S. K. Scott, J. Cogn. Neurosci. 28, 483–500 (2016).

40. S. K. Scott, S. Rosen, L. Wickham, R. J. S. Wise, J. Acoust. Soc. Am. 115, 813–821 (2004).

41. S. K. Scott, S. Rosen, C. P. Beaman, J. Davis, R. J. S. Wise, J. Acoust. Soc. Am. 125, 1737–1743 (2009).

42. S. K. Scott, C. McGettigan, Hear. Res. 303, 58–66 (2013). 43. F. Schneider et al., Sci. Rep. 8, 17948 (2018).

44. C. I. Petkov et al., Nat. Neurosci. 11, 367–374 (2008).

45. J. S. Kyong et al., J. Cogn. Neurosci. 26, 1748–1763 (2014).

46. D. Sammler, M. H. Grosbras, A. Anwander, P. E. Bestelmeyer, P. Belin, Curr. Biol. 25, 3079–3085 (2015).

47. S. R. Mathias, K. von Kriegstein, Front. Biosci. (Schol. Ed.) 6, 92–109 (2014).

48. S. C. Creel, M. R. Bregman, Lang. Linguist. Compass 5, 190–204 (2011).

49. Z. G. Cai et al., Cognit. Psychol. 98, 73–101 (2017).

50. F. Eisner, J. M. McQueen, Percept. Psychophys. 67, 224–238 (2005).

51. F. Eisner, J. M. McQueen, J. Acoust. Soc. Am. 119, 1950–1953 (2006).

52. T. K. Perrachione, S. N. Del Tufo, J. D. Gabrieli, Science 333, 595 (2011).

53. J. Y. Choi, T. K. Perrachione, Brain Lang. 196, 104655 (2019).

54. R. J. S. Wise et al., Brain 124, 83–95 (2001).

55. G. Hickok, B. Buchsbaum, C. Humphries, T. Muftuler, J. Cogn. Neurosci. 15, 673–682 (2003).

56. Z. K. Agnew, C. McGettigan, B. Banks, S. K. Scott, Neuroimage 73, 191–199 (2013).

57. H. Takaso, F. Eisner, R. J. S. Wise, S. K. Scott, Journal of Speech Hearing and Language Research. 53, 226–236 (2010).

58. K. M. Jasmin et al., J. Neurosci. 36, 4669–4680 (2016).

59. C. McGettigan et al., J. Cogn. Neurosci. 25, 1875–1886 (2013).

60. H. Blank, M. H. Davis, PLOS Biol. 14, e1002577 (2016).

61. Y. Sun et al., Brain Lang. 149, 55–65 (2015).

62. C. McGettigan et al., J. Cogn. Neurosci. 23, 961–977 (2011).

63. A. Keitel, J. Gross, C. Kayser, PLOS Biol. 16, e2004473 (2018).

64. Y. Oganian, E. G. Chang, BioRxiv 388280 [Preprint]. 9 August 2018. https://doi.org/10.1101/388280.

65. L. M. Hyman, Phonology 28, 55–85 (2011).

66. C. Jacquemot, C. Pallier, D. LeBihan, S. Dehaene, E. Dupoux, J. Neurosci. 23, 9541–9546 (2003).

67. K. Jasmin, C. F. Lima, S. K. Scott, Nat. Rev. Neurosci. 20, 425–434 (2019).

68. L. S. Hamilton, E. Edwards, E. F. Chang, Curr. Biol. 28, 1860–1871.e4 (2018).

69. S. Norman-Haignere, N. G. Kanwisher, J. H. McDermott, Neuron 88, 1281–1296 (2015).

70. K. Fiehler, I. Schütz, T. Meller, L. Thaler, Multisens. Res. 28, 195–226 (2015).
71. A. Callan, D. E. Callan, H. Ando, Neuroimage 66, 22–27 (2013).
72. L. Ceravolo, S. Frühholz, D. Grandjean, Soc. Cogn. Affect. Neurosci. 11, 793–802 (2016).
73. S. Garrod, M. J. Pickering, Top. Cogn. Sci. 1, 292–304 (2009).
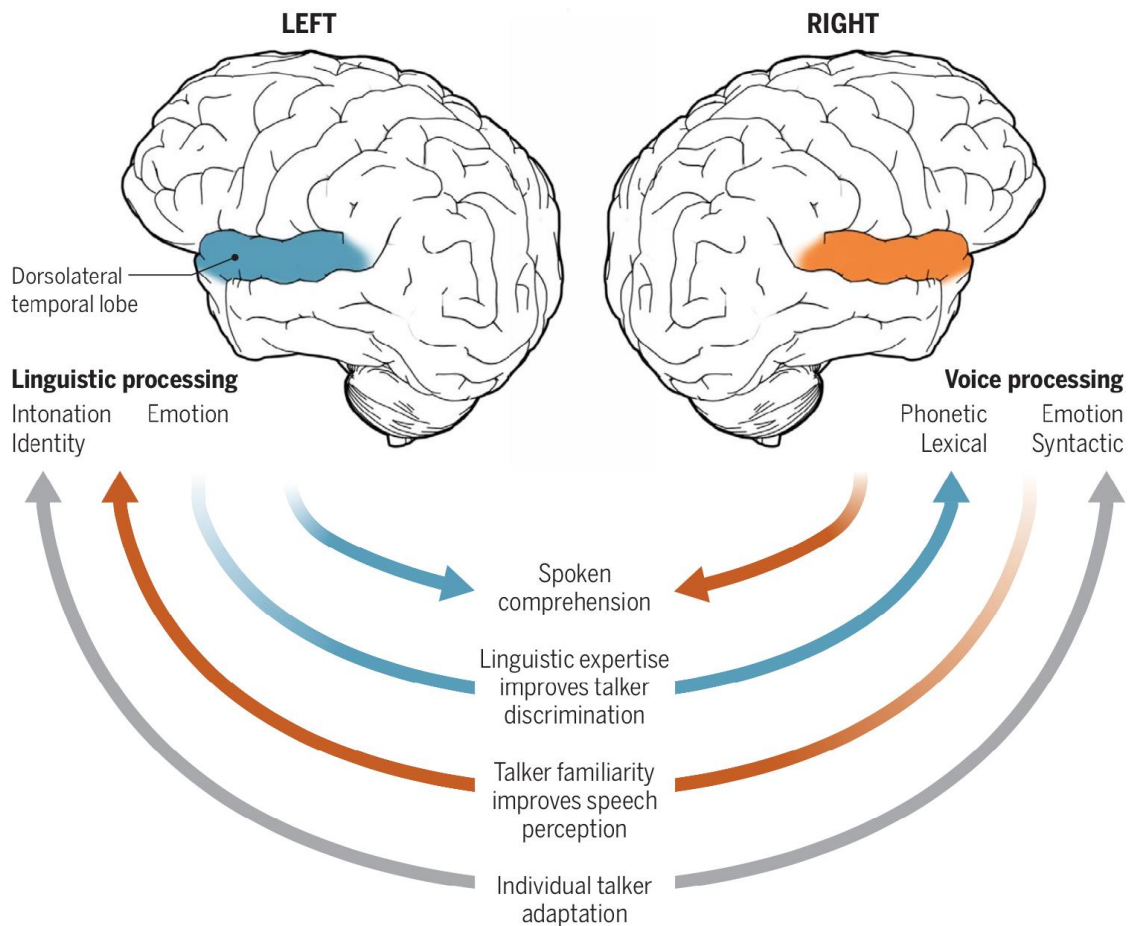74. K. Jasmin et al., Brain 142, 808–822 (2019).

Figure 1

Fig. 1. Schematic representation of the left and right dorsolateral temporal lobes. Shown are the left (yellow) and right (green) dorsolateral temporal lobes and the different candidate interactions that may occur between linguistic and nonlinguistic aspects of voice and language recognition processes on the basis of behavioral and noninvasive stimulation studies.

Figure 2

Perceptually relevant speech tracking at the **phrase level (0.6–1.3 Hz),** seen in the left precentral (1) and postcentral (2) gyri, supramarginal gyrus (3), and Heschl's gyrus (not shown) with the peak in the left premotor cortex (5).
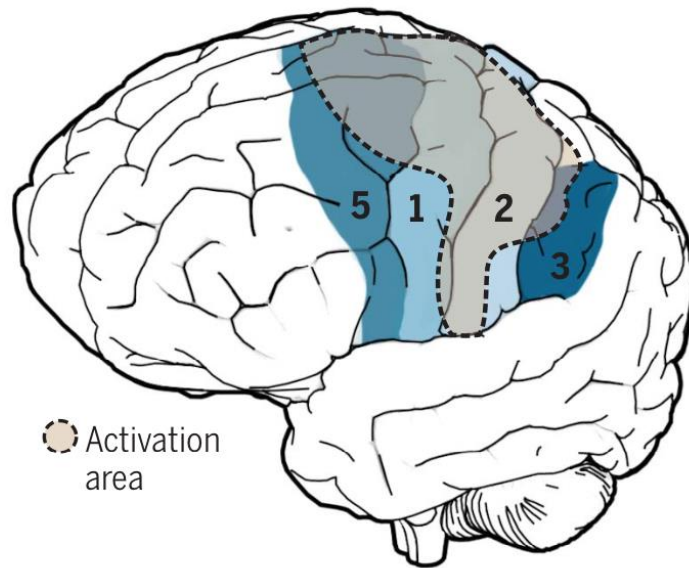
Perceptually relevant speech tracking at the **word level (1.8–3.0 Hz),** seen in the left superior (1), middle (2), and inferior temporal gyrus (3), supramarginal gyrus (4) and Heschl's gyrus (not shown). The effect peaked in the left middle temporal gyrus.
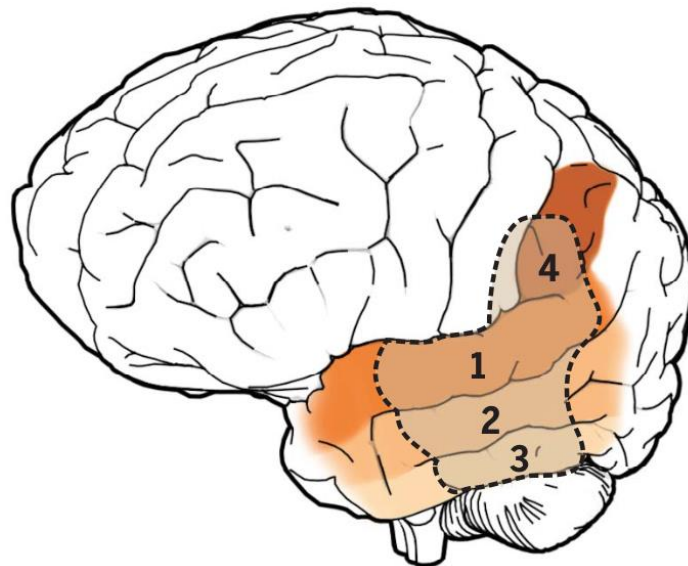

Fig. 2. Summary of some of the findings from Keitel et al. In Keitel *et al.* (*63*), naturalistic speech was played to listeners who were scanned with magnetoencephalography (MEG). The authors then interrogated the cortical activity based on the statistical properties of the speech from the original stimuli. They found that perceptually relevant speech tracking (based on mutual information, a measure of synchrony in MEG studies) was larger for correct trials (relative to incorrect trials) in premotor and motor cortex for activity at the time scale of phrases (0.6 to 1.3 Hz) (top) and in the left temporal lobe for activity at the time scale of words (1.8 to 3 Hz) (bottom). Activity is shown schematically within the dotted fields.
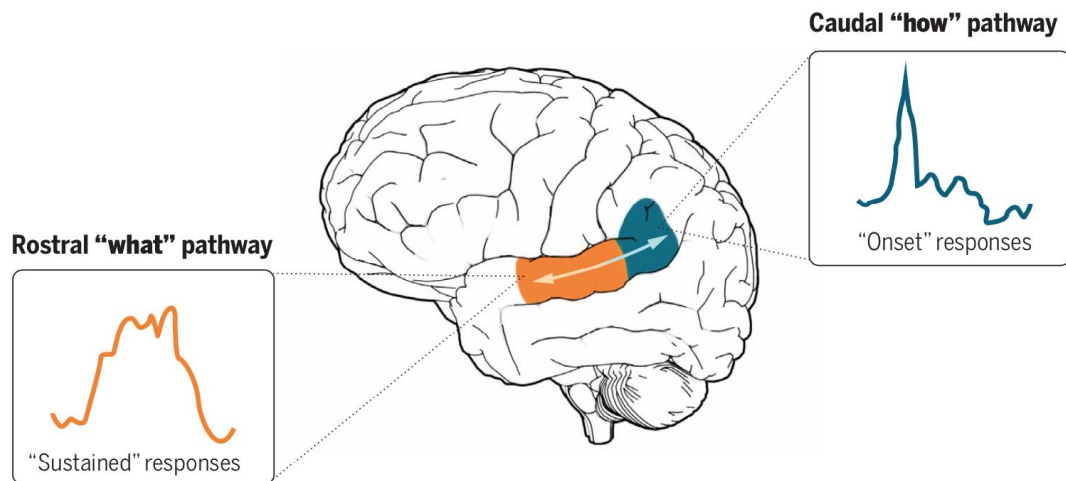
Figure 3

Fig. 3. Schematic representation of the difference between slow, sustained and fast, transient neural responses found in rostral (yellow) and caudal (blue) pathways, respectively. These differences were established by categorizing the time course of eCOG responses to all stimuli presented to participants across a range of studies (*68*). This difference in temporal response properties is one of several new findings about rostral-caudal auditory differences in responses to sounds that have been incorporated into a new domain-general theory of auditory processing (*67*). Shown is the difference between a yellow rostral "what" pathway and a blue caudal "how" pathway and includes exemplar profiles.