



The neuroanatomical and functional organization of speech perception

Sophie K. Scott¹ and Ingrid S. Johnsrude²

¹Department of Psychology, University College London, Gower Street, London, UK WC1E 6BT.

²MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge, UK CB2 2EF.

A striking property of speech perception is its resilience in the face of acoustic variability (among speech sounds produced by different speakers at different times, for example). The robustness of speech perception might, in part, result from multiple, complementary representations of the input, which operate in both acoustic–phonetic feature-based and articulatory–gestural domains. Recent studies of the anatomical and functional organization of the non-human primate auditory cortical system point to multiple, parallel, hierarchically organized processing pathways that involve the temporal, parietal and frontal cortices. Functional neuroimaging evidence indicates that a similar organization might underlie speech perception in humans. These parallel, hierarchical processing ‘streams’, both within and across hemispheres, might operate on distinguishable, complementary types of representations and subserve complementary types of processing. Two long-opposing views of speech perception have posited a basis either in acoustic feature processing or in gestural motor processing; the view put forward here might help reconcile these positions.

In speech perception, a stream of sound is decoded into a meaningful utterance, apparently involuntarily. The apparent automaticity belies the many stages of processing that must be invoked to extract speech-relevant information from the stream of sound, to map this onto stored representations and to combine these representations to derive the overall meaning. Recent studies in macaque monkeys, as well as in humans, provide information about the processing stages required, the representations that might be used and the neural bases of speech perception. Anatomical and neurophysiological findings in monkeys support the idea of multiple, parallel streams of processing in the primate auditory system [1] (Box 1 reviews recent proposals concerning the functional specialization of different auditory ‘streams’). Results of functional neuroimaging studies provide evidence that human speech perception might indeed be based on multiple, hierarchical processing pathways, and that different kinds of representations could be preferentially treated in these different ‘streams’ (such as acoustic–phonetic features and articulatory gestures). Here, we review recent work relevant to the neurobiological

implementation of speech perception. The evidence suggests: (1) that existing concepts of functional specialization for language in the brain need to be refined to accommodate evidence of anatomical and functional heterogeneity within language areas; (2) that there are multiple parallel and hierarchical processing systems for speech perception in the brain; and (3) that these might use fundamentally different representations. We consider speech as conceptually distinct from language, which, with its syntactic and generative aspects and highly elaborated semantic structure, is likely to be unique to humans. By contrast, the brain organization that supports the perception of speech sounds as communicative acts is probably not unique, and this phylogenetically older system must constrain the way in which all aspects of language are neurally organized and represented [2].

Anatomical and functional organization of the auditory cortex

The incoming speech signal is probably highly processed and re-coded by the time it reaches the auditory cortex, given the complexity of the ascending auditory pathway [3,4]. Although there is considerable ‘tuning’ in the auditory system to the acoustic properties of speech, the processing operations conducted in the relay nuclei of the brainstem and thalamus are general to all sounds, and speech-specific operations probably do not begin until the signal reaches the cerebral cortex. The anatomical and functional organization of the auditory cortex is still not well understood, particularly in humans. Microelectrode recordings can be undertaken only in rare circumstances (e.g. during neurosurgery) [5], postmortem histological material is difficult to obtain [6–9] and *in vivo* tracer studies in humans are currently not possible. Therefore, the anatomical and functional organization of the human auditory cortical system has to be largely inferred for studies in non-human primates (Fig. 1). Primate cortical auditory areas are conceptually organized into a ‘core’ of primary-like auditory cortices (recipients of thalamic input from the ventral medial geniculate complex, predominantly) that project to a surrounding ‘belt’, which is composed of several different cortical areas [10–18]. These core and belt areas display some functional specificity: for example, tonotopy is seen in the core regions that respond to pure tones, and differential responses to noise bandwidths are seen running laterally across the belt region [19]. Belt areas connect with lateral ‘parabelt’

Corresponding authors: Sophie K. Scott (sophie.scott@ucl.ac.uk), Ingrid S. Johnsrude (ingrid.johnsrude@mrc-cbu.cam.ac.uk).

Box 1. Multiple streams of processing in auditory cortices and the speech perception system

The concept of multiple, parallel processing streams is well established in other areas of cognitive neuroscience such as memory [a] and vision [b]. In the visual domain, this classically includes a 'what' pathway, which is crucial for object identification and extends anteriorly from primary visual areas through ventral parietal and temporal-lobe regions, and a 'where' pathway, which is crucial for object localization and extends from primary visual areas through posterior parietal cortex [b]. This view has subsequently been modified, such that the 'where' pathway is also considered to be a 'how' pathway, which is important for the determination of egocentric object position and preparation for visually guided actions (e.g. grasping) [c].

The framework of segregation of auditory anatomy into multiple, parallel processing streams has been elaborated in the functional domain with evidence from electrophysiological studies in primates. Such studies indicate a functional dissociation between anterior and posterior streams. A 'what' stream associated with vocalizations (presumably subserving auditory object identification) involves the anterior belt and parabelt, and anterior temporal and ventrolateral frontal regions (analogous to the visual ventral pathway); a 'where' stream for sound localization involves the posterior belt and parabelt, and posterior temporal and dorsolateral frontal regions (analogous to the visual dorsal pathway) [d–f]. (The middle zone of belt and parabelt, which is between the anterior and posterior divisions, makes frontal connections that overlap with both of the other two, suggesting the possibility of a third stream of intermediate function [g].) Posterior auditory regions have been shown to respond to spatial cues in primate studies [f]. There is also considerable functional imaging data that supports the existence of posterior temporal–parietal involvement in the processing of auditory spatial information (i.e. a posterior 'where' system) [h–j], although the degree to which this might be purely spatial has been disputed [k]. There is also functional imaging evidence for an anterior 'what' stream in human speech processing [l].

Zatorre and Belin [m] suggest that the 'dorsal' route of processing, which is posterior to the primary auditory cortex (PAC) (a 'how' pathway), might subservise perception of the evolution of a sound in time – its spectral dynamics. This, in the context of vocalizations, would correspond to changes in formant energy over time, which would reflect the movement of the vocal apparatus. They further suggest that the meaning of an utterance is extracted via this dorsal route, whereas the 'ventral' path (which runs anterior to PAC) allows identification of a sound source (i.e. the speaker).

In a model of language processing, Hickok and Poeppel [n] place prelexical processing of the speech signal in auditory cortices bilaterally, and an interface between sound and meaning (which is

assumed to be cortically widely distributed) in the posterior temporal–parietal–occipital junction (primarily in the left hemisphere). They refer to this as a 'ventral stream', although it does not extend anteriorly along the temporal lobe as in other schemes. The dorsal pathway in their model, which comprises inferior parietal and frontal cortices, is suggested to be important for auditory–motor integration. Binder *et al.* [o] have also formulated a model of the mapping from sound to meaning, which runs lateral and ventral to auditory cortex, with semantic representations dependent upon inferior temporal gyrus.

References

- a Squire, L.R. (1992) Memory and the hippocampus: a synthesis from findings with rats, monkeys, and humans. *Psychol. Rev.* 99, 195–231
- b Ungerleider, L.G. and Mishkin, M. (1982) Two cortical visual systems. In *Analysis of Visual Behavior* (Ingle, D.J. *et al.*, eds), pp. 549–586, MIT Press
- c Goodale, M.A. and Milner, A.D. (1992) Separate visual pathways for perception and action. *Trends Neurosci.* 15, 20–25
- d Recanzone, G.H. (2001) Spatial processing in the primate auditory cortex. *Audiol. Neurootol.* 6, 178–181
- e Romanski, L.M. *et al.* (1999) Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nat. Neurosci.* 2, 1131–1136
- f Tian, B. *et al.* (2001) Functional specialization in rhesus monkey auditory cortex. *Science* 292, 290–293
- g Kaas, J.H. and Hackett, T.A. (1999) 'What' and 'where' processing in auditory cortex. *Nat. Neurosci.* 2, 1045–1047
- h Griffiths, T.D. *et al.* (1998) Right parietal cortex is involved in the perception of sound movement in humans. *Nat. Neurosci.* 1, 74–79
- i Warren, J.D. *et al.* (2002) Perception of sound-source motion by the human brain. *Neuron* 34, 139–148
- j Zatorre, R.J. *et al.* (2002) Where is 'where' in the human auditory cortex? *Nat. Neurosci.* 5, 905–909
- k Middlebrooks, J.C. (2002) Auditory space processing: here, there or everywhere? *Nat. Neurosci.* 5, 824–826
- l Scott, S.K. *et al.* (2000) Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* 123, 2400–2406
- m Belin, P. and Zatorre, R.J. (2000) 'What', 'where' and 'how' in auditory cortex. *Nat. Neurosci.* 3, 965–966
- n Hickok, G. and Poeppel, D. (2000) Towards a functional neuro-anatomy of speech perception. *Trends Cogn. Sci.* 4, 131–138
- o Binder, J.R. *et al.* (2000) Human temporal lobe activation by speech and nonspeech sounds. *Cereb. Cortex* 10, 512–528

fields, and both belt and parabelt fields project to the temporal, parietal and frontal lobes. Each of these regions – core, belt and parabelt – has a distinctive histological composition, specific thalamocortical and corticocortical connections, and unique physiological and functional characteristics [6,12,13,15,16,18–22]. The core, belt and parabelt areas are strikingly hierarchical in their connections and response properties, suggesting that at least three discrete levels of processing within the auditory region can be supported [13,15–20]. Studies of human auditory anatomy and physiology, although few in number relative to the work in macaques, reveal a similar anatomical organization. In humans, the first transverse gyrus of Heschl (HG) contains cytoarchitecturally identifiable 'core' auditory cortex.

Functional imaging studies, using positron emission tomography (PET) or functional magnetic resonance imaging (fMRI), reveal activation of HG when listening to any type of sound when compared with silence, or

when the amount of stimulation is parametrically manipulated [23–25]. Activation of HG is not specific to speech, and the response properties of this tissue, revealed using functional neuroimaging, are consistent with electrophysiological studies of core regions in other animals (e.g. tonotopic gradients are observed [26,27]).

Multiple processing 'streams' anterior and ventral to the auditory cortex

Projections of belt and parabelt areas beyond the auditory region are topographically organized. In the macaque monkey, anterior belt and parabelt areas connect with different regions within the anterior superior temporal gyrus (STG) and sulcus (STS) [13,16,28], and with multiple sites within the orbitofrontal, ventrolateral and dorsolateral frontal cortices, including areas 46, 12 and 45 [29–32]. A recent study in macaques provides electrophysiological evidence for an auditory-specific area in the

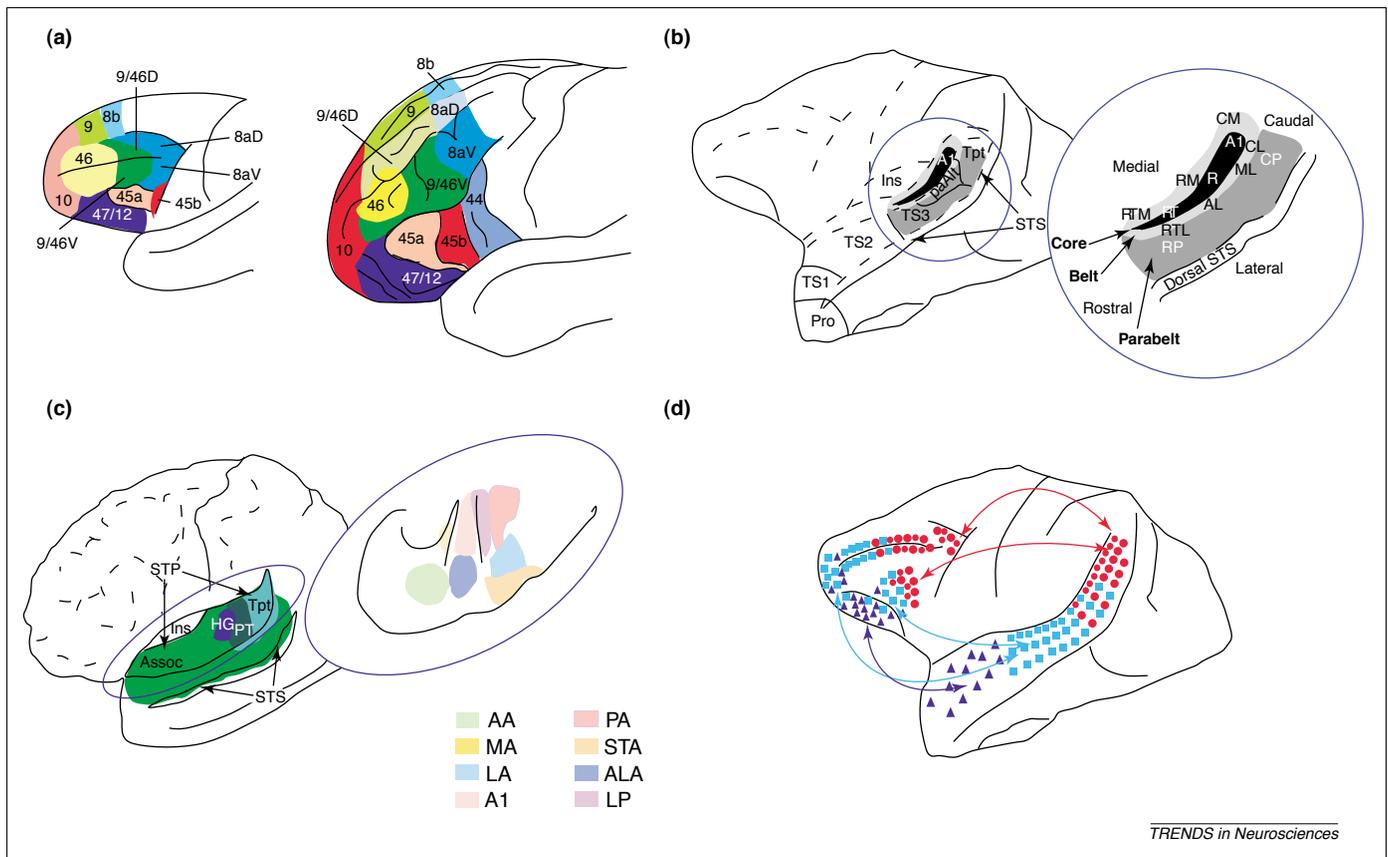


Fig. 1. Regions and connections of primate and human prefrontal and auditory cortex. (a) Specialization in the frontal cortex of the primate (left) and human (right). Text indicates area designation. Adapted from Ref. [94]. (b) Specialization in the primate temporal lobe (colour coded). Inset shows the supra-temporal plane in greater detail. (c) The lateral aspect of human brain, showing detail of the supra-temporal plane. Adapted from Ref. [32]. (d) Auditory-frontal projections. Caudal belt and parabelt regions and their reciprocal frontal connections are shown as red arrows and circles; rostral belt and parabelt regions and their projections are shown as pale blue arrows and squares; and anterior temporal-lobe regions and connections are shown as dark blue arrows and triangles. Larger squares and circles represent the contributions from the parabelt: these are greater than those of the belt. Adapted from Refs [9,40]. Abbreviations: A1, primary auditory cortex, auditory area 1; AA, anterior area; AL, anterior lateral belt; ALA, anterior lateral area; Assoc, association cortex; CL, caudal lateral belt; CM, caudal medial belt; CP, caudal auditory parabelt; HG, Heschl's gyrus; Ins, Insula; LA, lateral area; LP, lateroposterior area; MA, medial area; ML, medial-lateral belt; PA, posterior area; paAlt lateral parakoniocortex; PP, planum polare; Pro, proisocortex; PT, planum temporale; R, rostral area (core auditory cortex); RM, rostro-medial area; RP, rostral parabelt; RTL, lateral rostrotemporal auditory belt; RTM, medial rostrotemporal auditory belt; sts, superior temporal sulcus; STA, superior temporal area; STP, superior temporal plane; Tpt, temporal-parietal area; TS1, superior temporal area 1; TS2, superior temporal area 2; TS3, superior temporal area 3.

ventrolateral prefrontal cortex, which overlaps areas 12 and 45 [33]. Another recent study [34] suggests that much of what is referred to as area 12 in these studies is in fact area 45; importantly, the human homologue of area 45 is found in the inferior frontal gyrus (pars triangularis) and is considered to be one of the architectonic constituents of Broca's area [35] (Fig. 1).

Functional imaging studies of the perception of speech and other complex sounds in humans provide evidence for a hierarchy of processing that extends outwards from core regions, probably through the human equivalent of belt, parabelt and beyond (Box 2). The regions immediately anterolateral to the core appear selective for sounds with spectrotemporal structure, such as harmonic complexes and frequency-modulated sounds [36], and broadband noise that incorporates sufficient temporal regularity to yield a pitch percept (regular-interval noise) [37]. Responses to amplitude-modulated noise are also observed in this approximate location [38], as are responses to white noise co-modulated with the original speech signal [24,39,40]. Thus, complex spectrotemporal structure, including modulation of frequency

(FM) or amplitude (AM), leads to a specific pattern of response in what is probably anterolateral belt or parabelt [9,36,41].

Lateral to the HG on the convexity of the STG that extends into the STS, in what could be the human equivalent of parabelt cortex, activation is observed in response to stimuli with the acoustic features of phonetic cues [42] and to intelligible speech [43], as well as to harmonic tones, frequency-modulated tones [36] and sounds with changing spectral structure [23]. Thus, although speech-related activation can be seen in this area, it cannot at present be spatially distinguished from responses to non-speech sounds with spectral detail and variation. This could suggest a degree of parallel processing of the speech input.

Activation specific to intelligible speech is observed in the left anterior STS at what could be, judging from the non-human primate anatomical literature, a processing level beyond those intrinsic to auditory cortices [43,44]. This region is multimodal, receiving projections from auditory, visual and somatosensory cortex in primates [28], and may be important in representing or accessing

Box 2. Relationship between structure and function in human auditory cortex

We present the results of a meta-analysis of the peaks of activation seen in functional imaging studies that have looked at speech and non-speech auditory processing in the temporal cortex. The analysis was limited to studies in which the properties of sound revealed by the contrast could be determined relative to the baselines used and, thus, is not a comprehensive review of functional imaging studies of language. Fig. 1 shows the superior surface of the left temporal lobe in a transaxial view, with the *x* and *y* coordinates (describing lateral–medial and anterior–posterior positions, respectively) of the peak activities plotted in standardized stereotaxic space. The location of the primary auditory cortex (PAC) is indicated schematically by the black line with grey circles at either end [a]. The coloured regions indicate a subset of the human auditory areas identified by Rivier and Clarke [b], which were described for the same stereotaxic space and are similar to those described by Wallace *et al.* [c]. Given the appropriate provisos that these studies compare different subjects, imaging techniques and analysis methods (as well as the use of peaks), and the fact that the data are collapsed in the *z* plane (dorsal–ventral), a pattern of activation can be seen in which a

hierarchy of processing and some relation to the Rivier and Clarke [b] auditory areas are evident – particularly for the primary and lateral auditory areas, which run anterolateral to PAC. A study that contrasted sound with silence (black squares) has shown activation in the core of auditory cortex, as well as other lateral regions [d]. The contrast of harmonically structured sounds with simple sine tones [e] (blue squares) has shown that activation lies anterolateral to the PAC. Responses to amplitude-varying stimuli are also seen in this anterolateral ‘parabelt’ region – the anterior lateral area (ALA) and lateral area (LA) in humans [b,c] – both to amplitude-modulated (AM) noise over unmodulated noise [f] and to increasing rates of speech-envelope-modulated noise [signal-correlated noise (SCN)] [g] (green squares). Frequency-modulated (FM) tones contrasted with unchanging tones (purple triangles) activate lateral parabelt regions [e] and also more anterior and posterior superior temporal regions, extending down into the superior temporal sulcus (STS). Spectrally dynamic sounds, relative to unchanging sounds (purple circles), activate the anterolateral parabelt and the more anterior superior temporal gyrus (STG) [d] (including the anterior area AA [b,c]). Speech-related stimuli, such as those with phonetic cues and features regardless of intelligibility, activate posterior and anterior lateral superior temporal regions (pink circles) [h], whereas intelligible speech against baselines of equal acoustic structural complexity (red squares) activates anterior STG and STS [h].

There is, thus, some evidence for parallel and hierarchical responses in human auditory areas. There are peaks in ‘core’ PAC in responses to sounds against silence; in an anterolateral region possibly comprising belt and parabelt cortex in primates (and relating to ALA and LA auditory areas in humans), there are responses to spectro-temporal structure in the sound. Lateral and anterior to this there are responses to stimuli with phonetic cues and features; intelligible speech responses lie anterior and ventral to this, consistent with an anterior ‘what’ processing stream for mapping from the acoustic signal onto speech. Note that there are peak responses to sounds lying posterior to the ‘core’ PAC, in the planum temporale and more ventrally in the STS, consistent with other sound-related processing pathways.

References

- a Penhune, V.B. *et al.* (1996) Interhemispheric anatomical differences in human primary auditory cortex: probabilistic mapping and volume measurement from magnetic resonance scans. *Cereb. Cortex* 6, 661–672
- b Rivier, F. and Clarke, S. (1997) Cytochrome oxidase, acetylcholinesterase, and NADPH-diaphorase staining in human supratemporal and insular cortex. *NeuroImage* 6, 288–304
- c Wallace, M.N. *et al.* (2002) Histochemical identification of cortical areas in the auditory region of the human brain. *Exp. Brain Res.* 143, 499–508
- d Thivard, L. *et al.* (2000) A cortical region sensitive to auditory spectral motion. *NeuroReport* 11, 2969–2972
- e Hall, D. *et al.* (2002) Spectral and temporal processing in human auditory cortex. *Cereb. Cortex* 12, 140–149
- f Giraud, A.L. *et al.* (2000) Representation of the temporal envelope of sounds in the human brain. *J. Neurophysiol.* 84, 1588–1598
- g Mummery, C.J. *et al.* (1999) Functional neuroimaging of speech perception in six normal and two aphasic subjects. *J. Acoust. Soc. Am.* 106, 449–457
- h Scott, S.K. *et al.* (2000) Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* 123, 2400–2406

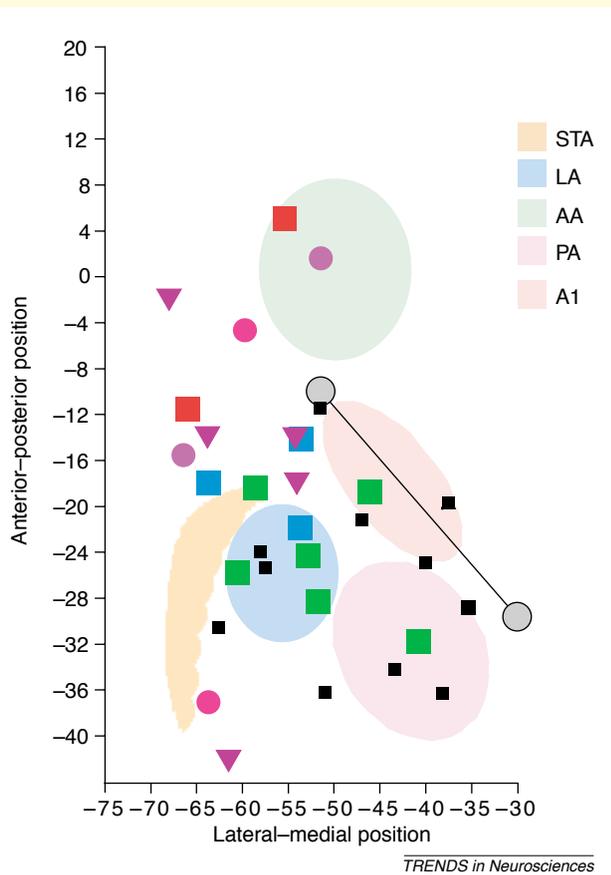


Fig. 1. Peak activations of neural responses to different acoustic contrasts on the human left dorsal temporal lobe, revealed by functional imaging. Abbreviations: A1, primary auditory cortex; AA, anterior area; LA, lateral area; PA, posterior area; STA, superior temporal area.

the meaning content of utterances. Interestingly, anatomical studies in macaques indicate that both rostral and caudal parabelt project to most of the STS, in an overlapping fashion. Only the anteriormost STS appears to receive projections primarily from rostral parabelt [13]. In humans, support for a role in semantics (in accessing word meanings or representing the lexical and semantic

information) for the anterior temporal-lobe region comes from patients with semantic dementia, who exhibit a progressive deterioration in the comprehension of single words, with preserved syntax and object use [45]. This disorder shows a characteristic neural profile of grey matter loss in the left anterior temporal lobe [46,47], with sparing of white matter tracts. The anterior temporal

cortex is heavily interconnected with the ventrolateral prefrontal cortex in humans, as it is in monkeys [48,49], providing a route for frontal-lobe systems to have access to word-meaning information. Activation in both the anterior temporal and inferior frontal cortex is evident in functional neuroimaging studies when an active task is performed upon the speech input [44,50], and this can be extended to examples of explicit semantic processing of linguistic input [51–53].

As the disorder of semantic dementia progresses, atrophy is also observed posteriorly and ventrally in the left temporal lobe, and deterioration can also be observed on the right. Later stages of the disease thus compromise function in what has been called the basal language area, in the ventral temporal cortex. Functional imaging studies have shown responses to meaningful stimuli (presented via visual, auditory and tactile modalities) in the ventral temporal cortices, consistent with this area also being important for semantic representations [54–56].

Multiple processing ‘streams’ posterior and dorsal to the auditory cortex

The temporoparietal junction, which forms the interface between auditory cortex and parietal and frontal systems, is anatomically heterogeneous. A recent study in humans indicates at least four distinguishable areas on the anterior planum temporale alone [9]. For parcellation of the rest of this diverse region, including the posterior planum temporale (PT), parietal operculum and lateral superior, middle temporal and supramarginal gyri, we are reliant upon classical cyto- and myeloarchitectonic studies [57]. These studies suggest a further seven anatomically distinguishable areas, at least, and the refinement available with modern histochemical techniques will probably reveal further subdivisions within these areas. This anatomical heterogeneity probably reflects considerable functional specialisation. Griffiths and Warren [58] have recently suggested a model of PT function that highlights the anatomical and functional heterogeneity of this area. At the anterolateral edge of the planum temporale, activation is seen in response to both speech and nonspeech sounds with no clear differentiation [40,42,59,60] (Box 2), whereas a left-sided region posterior to the core in the medial PT [57] is activated during articulation, whether the subject speaks aloud or not [40,61–63]. Areas of the PT (particularly medial ones) and the parietal operculum also appear to be recruited during the processing of sound location and sound motion; the network can also be demonstrated to involve more superior parietal and conjugate frontal areas [64–66].

Posterior belt and parabelt areas connect with the STG and STS. The posterior STS also receives visual afferent input [28,67]. Activation in response to sound in the posterior STS region is not speech specific and it shows sensitivity to structure in visual sequences [68], as well as being involved in speech tasks that require auditory and visual integration, such as lip reading [69,70] and verbal fluency tasks [40]. Nonetheless, the left posterior STS is clearly important for speech function. Studies of patients with speech comprehension

deficits reveal consistent damage to the posterior STS [71]. Lesions restricted to left posterior STS can produce conduction aphasia: a condition in which patients can understand speech but not repeat it [72,73]. Repetition is an important aspect of speech development during both childhood and adulthood. Repetition can exist without access to word meaning: we can repeat non-words and foreign words (albeit constrained by our linguistic experience). Repetition of unfamiliar words (or non-words) probably relies on the ability to represent a sequence of sounds over a short time period, potentially ‘buffering’ the input; the left posterior STS might be important in this regard [40]. The transient representation of speech sounds with respect to producing them is a crucial aspect of the acquisition of language; new words are generally learnt via some overt articulation [74].

Posterior belt and parabelt areas (and the temporal-lobe areas to which they project, such as the STS) connect with multiple sites within dorsolateral frontal cortex in macaques. These include area 46 (which is heavily implicated in working memory), area 8a (which contains the frontal eye fields) and possibly area 6 (premotor cortex) [48]. Connections from posterior auditory areas tend to be more dorsolateral within prefrontal cortex than are projections from anterior auditory areas, but there can also be convergence of anterior and posterior auditory projections within the frontal lobe (Fig. 1). For example, both anterior and posterior non-primary auditory regions project to area 46, although perhaps to different subdivisions [30–32,48]. Both areas 46 and 8 have substantial connections with the premotor cortex [75,76]. Belt and/or parabelt also project to the caudate and putamen (part of the basal ganglia) in a topographic fashion [77]. These connections could provide a substrate for an auditory–motor system.

Production and perception of speech are very likely to be functionally intertwined, and several models of speech perception, positing a basis in articulatory or gestural representations, have been formulated [78–81]. The motor theory of speech perception, for example, takes the strong stance that the direct perception of gestures underlies speech processing, with minimal decoding of the acoustic input. More moderate positions that emphasize auditory processing also accept that speech perception and production are linked [82], and this association is seen most clearly in the development of language [83,84].

A system for observation–execution matching has been described in primates. Ventral premotor cortex, possibly including the primate homologue of Broca’s area, contains neurons that discharge not only when a monkey grasps or manipulates objects, but also when it observes similar actions performed by others. Such neurons are called ‘mirror neurons’ [81,85]. Furthermore, neurons in this part of the brain appear to respond similarly to mouth actions and manual gestures, and also to respond to the sounds produced by those manual gestures (e.g. the sound of ripping paper) [86,87]. This could be evidence for an auditory–vocal

(perception–production) homologue to the visual–haptic observation–execution mirror neuron system [81,85]. Consistent with this work, recent studies in humans have revealed that speech perception, either watching speaking mouths or listening to speech without visual input, enhances the excitability of motor cortex in the left hemisphere relative to nonlinguistic baseline perceptual tasks [88,89]. Although it remains to be seen whether this finding is truly specific to speech, it is consistent with the idea that speech perception involves a system that provides an excitatory input to the motor face area, perhaps via motor (articulatory–gestural) representations. Perhaps mirror neurons need not be activated solely by visual input; gestures might be represented in the acoustic signal and in stored motor knowledge, and could be activated when processing the incoming sound.

Concluding remarks

A functional distinction can, thus, be made between streams of processing that extend anteriorly and inferolaterally from primary auditory core. Anteriorly, these involve anterior belt and parabelt, the polymodal cortex of the anterior STS, and ventrolateral and dorsolateral frontal cortex. Posterior routes involve posterior auditory belt and parabelt, polymodal cortex of the posterior STS, parietal cortex, and ventrolateral and dorsolateral frontal cortex. We have presented evidence that the anterior system might be important for mapping acoustic–phonetic cues onto lexical representations, whereas the posterior system might process articulatory–gestural representations of speech acts. This could resolve the longstanding debate between accounts of speech perception that favour acoustic feature processes and those that emphasize gestural motor processing. Motor and acoustic–phonetic representations will, typically, operate in a highly inter-related fashion; thus, the silent mouthing of a word primes a later auditory lexical decision but not a visual lexical decision [90]. There is also the potential for integration with the theories of spatial processing in audition (Box 1), as psychophysical studies indicate that sounds can be heard only as a single vowel if they can be grouped together by location [91].

In terms of the classical neuroanatomy of speech, this account has the benefit of expanding the roles of Broca's and Wernicke's areas as output and input systems, and differentiating subsystems within these. Broca's area (comprising prefrontal and premotor cortices) is involved in some speech processing tasks, including explicit speech sound segmentation [92], whereas 'Wernicke's area' (comprising a large and anatomically heterogeneous portion of temporal–parietal cortex) encompasses both speech recognition systems and representations involved in the output of speech [40].

As in the visual system, therefore [93], the relationship between a hierarchically organized perceptual system and functionally separable processing streams can be described in the auditory system, with a consequent impact on the processing of human speech sounds. Speech is processed both as a sound and as an action. Further

work will determine how this affects specialisation (including hemispheric lateralization) for speech in the human brain.

Acknowledgements

S.K.S. is supported by the Wellcome Trust and I.S.J. by the Medical Research Council. We wish to thank Matt Davis, Troy Hackett, Liz Romanski, Kate Watkins, Richard Wise and Ed Yeterian for helpful discussion and comments on earlier versions of this paper.

References

- Rauschecker, J.P. (1998) Cortical processing of complex sounds. *Curr. Opin. Neurobiol.* 8, 516–521
- Ghazanfar, A.A. and Hauser, M.D. (1999) The neuroethology of primate vocal communication: substrates for the evolution of speech. *Trends Cogn. Sci.* 3, 377–384
- Eggermont, J.J. (2001) Between sound and perception: reviewing the search for a neural code. *Hear. Res.* 157, 1–42
- Frisina, R.D. (2001) Subcortical neural coding mechanisms for auditory temporal processing. *Hear. Res.* 158, 1–27
- Howard, M.A. *et al.* (2000) Auditory cortex on the human posterior superior temporal gyrus. *J. Comp. Neurol.* 416, 79–92
- Hackett, T.A. *et al.* (2001) Architectonic identification of the core region in auditory cortex of macaques, chimpanzees, and humans. *J. Comp. Neurol.* 441, 197–222
- Morosan, P. *et al.* (2001) Human primary auditory cortex: cytoarchitectonic subdivisions and mapping into a spatial reference system. *NeuroImage* 13, 684–701
- Rademacher, J. *et al.* (2001) Probabilistic mapping and volume measurement of human primary auditory cortex. *NeuroImage* 13, 669–683
- Wallace, M.N. *et al.* (2002) Histochemical identification of cortical areas in the auditory region of the human brain. *Exp. Brain Res.* 143, 499–508
- Pandya, D.N. and Sanides, F. (1973) Architectonic parcellation of the temporal operculum in rhesus monkey and its projection pattern. *Z. Anat. Entwicklungsgesch.* 139, 127–161
- Merzenich, M.M. and Brugge, J.F. (1973) Representation of the cochlear partition of the superior temporal plane of the macaque monkey. *Brain Res.* 50, 275–296
- Hackett, T.A. *et al.* (1998) Thalamocortical connections of the parabelt auditory cortex in macaque monkeys. *J. Comp. Neurol.* 400, 271–286
- Hackett, T.A. *et al.* (1998) Subdivisions of auditory cortex and ipsilateral cortical connections of the parabelt auditory cortex in macaque monkeys. *J. Comp. Neurol.* 394, 475–495
- Jones, E.G. *et al.* (1995) Subdivisions of macaque monkey auditory cortex revealed by calcium-binding protein immunoreactivity. *J. Comp. Neurol.* 362, 153–170
- Kaas, J. and Hackett, T. (2000) Subdivisions of auditory cortex and processing streams in primates. *Proc. Natl. Acad. Sci. U. S. A.* 97, 11793–11799
- Pandya, D.N. (1995) Anatomy of the auditory cortex. *Rev. Neurol. (Paris)* 151, 486–494
- Rauschecker, J.P. (1998) Parallel processing in the auditory cortex of primates. *Audiol. Neurootol.* 3, 86–103
- 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. *et al.* (1995) Processing of complex sounds in the macaque nonprimary auditory cortex. *Science* 268, 111–114
- Kaas, J.H. *et al.* (1999) Auditory processing in primate cerebral cortex. *Curr. Opin. Neurobiol.* 9, 164–170
- Morel, A. *et al.* (1993) Tonotopic organization, architectonic fields, and connections of auditory cortex in macaque monkeys. *J. Comp. Neurol.* 335, 437–459
- Tian, B. *et al.* (2001) Functional specialization in rhesus monkey auditory cortex. *Science* 292, 290–293
- Thivard, L. *et al.* (2000) A cortical region sensitive to auditory spectral motion. *NeuroReport* 11, 2969–2972
- Mummery, C.J. *et al.* (1999) Functional neuroimaging of speech

- perception in six normal and two aphasic subjects. *J. Acoust. Soc. Am.* 106, 449–457
- 25 Johnsrude, I.S. *et al.* (2002) Functional imaging of the auditory system: the use of positron emission tomography. *Audiol. Neurootol.* 7, 251–276
- 26 Talavage, T.M. *et al.* (2000) Frequency-dependent responses exhibited by multiple regions in human auditory cortex. *Hear. Res.* 150, 225–244
- 27 Howard, M.A. III *et al.* (1996) A chronic microelectrode investigation of the tonotopic organization of human auditory cortex. *Brain Res.* 724, 260–264
- 28 Cusick, C.G. *et al.* (1995) Chemoarchitectonics and corticocortical terminations within the superior temporal sulcus of the rhesus monkey: evidence for subdivisions of superior temporal polysensory cortex. *J. Comp. Neurol.* 360, 513–535
- 29 Petrides, M. and Pandya, D.N. (1984) Projections to the frontal cortex from the posterior parietal region in the rhesus monkey. *J. Comp. Neurol.* 228, 105–116
- 30 Hackett, T. *et al.* (1999) Prefrontal connections of the parabelt auditory cortex in macaque monkeys. *Brain Res.* 817, 45–58
- 31 Romanski, L. *et al.* (1999) Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nat. Neurosci.* 2, 1131–1136
- 32 Romanski, L. *et al.* (1999) Auditory belt and parabelt projections to the prefrontal cortex in the rhesus monkey. *J. Comp. Neurol.* 403, 141–157
- 33 Romanski, L. and Goldman-Rakic, P. (2002) An auditory domain in primate prefrontal cortex. *Nat. Neurosci.* 5, 15–16
- 34 Petrides, M. and Pandya, D.N. (2002) Comparative cytoarchitectonic analysis of the human and the macaque ventrolateral prefrontal cortex and corticocortical connection patterns in the monkey. *Eur. J. Neurosci.* 16, 291–310
- 35 Amunts, K. *et al.* (1999) Broca's region revisited: cytoarchitecture and intersubject variability. *J. Comp. Neurol.* 412, 319–341
- 36 Hall, D. *et al.* (2002) Spectral and temporal processing in human auditory cortex. *Cereb. Cortex* 12, 140–149
- 37 Griffiths, T.D. *et al.* (2001) Encoding of the temporal regularity of sound in the human brainstem. *Nat. Neurosci.* 4, 633–637
- 38 Giraud, A.L. *et al.* (2000) Representation of the temporal envelope of sounds in the human brain. *J. Neurophysiol.* 84, 1588–1598
- 39 Schroeder, M.R. (1968) Reference signal for signal quality studies. *J. Acoust. Soc. Am.* 44, 1735–1736
- 40 Wise, R.J. *et al.* (2001) Separate neural subsystems within 'Wernicke's area'. *Brain* 124, 83–95
- 41 Wessinger, C. *et al.* (2001) Hierarchical organization of the human auditory cortex revealed by functional magnetic resonance imaging. *J. Cogn. Neurosci.* 13, 1–7
- 42 Binder, J.R. *et al.* (2000) Human temporal lobe activation by speech and nonspeech sounds. *Cereb. Cortex* 10, 512–528
- 43 Scott, S.K. *et al.* (2000) Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* 123, 2400–2406
- 44 Davis, M.H. and Johnsrude, I. Acoustic and linguistic processes in spoken language comprehension
- 45 Hodges, J. *et al.* (1999) 'What' and 'how': evidence for the dissociation of object knowledge and mechanical problem-solving skills in the human brain. *Proc. Natl. Acad. Sci. U. S. A.* 96, 9444–9448
- 46 Chan, D. *et al.* (2001) Patterns of temporal lobe atrophy in semantic dementia and Alzheimer's disease. *Ann. Neurol.* 49, 433–442
- 47 Mummery, C.J. *et al.* (2000) A voxel-based morphometry study of semantic dementia: relationship between temporal lobe atrophy and semantic memory. *Ann. Neurol.* 47, 36–45
- 48 Petrides, M. and Pandya, D.N. (1988) Association fiber pathways to the frontal cortex from the superior temporal region in the rhesus monkey. *J. Comp. Neurol.* 273, 52–66
- 49 Mamata, H. *et al.* (2002) High-resolution line scan diffusion tensor MR imaging of white matter fiber tract anatomy. *Am. J. Neuroradiol.* 23, 67–75
- 50 Poldrack, R.A. *et al.* (2001) Relations between the neural bases of dynamic auditory processing and phonological processing: evidence from fMRI. *J. Cogn. Neurosci.* 13, 687–697
- 51 Thompson-Schill, S.L. *et al.* (1997) Role of left inferior prefrontal cortex in retrieval of semantic knowledge: a reevaluation. *Proc. Natl. Acad. Sci. U. S. A.* 94, 14792–14797
- 52 Wagner, A.D. *et al.* (2001) Recovering meaning: left prefrontal cortex guides controlled semantic retrieval. *Neuron* 31, 329–338
- 53 Binder, J.R. *et al.* (1997) Human brain language areas identified by functional magnetic resonance imaging. *J. Neurosci.* 17, 353–362
- 54 Luders, H. *et al.* (1991) Basal temporal language area. *Brain* 114, 743–754
- 55 Buchel, C. *et al.* (1998) A multimodal language region in the ventral visual pathway. *Nature* 394, 274–277
- 56 Wise, R.J. *et al.* (2000) Noun imageability and the temporal lobes. *Neuropsychologia* 38, 985–994
- 57 Galaburda, A. and Sanides, F. (1980) Cytoarchitectonic organization of the human auditory cortex. *J. Comp. Neurol.* 190, 597–610
- 58 Griffiths, T.D. (2002) The planum temporale as a computational hub. *Trends Neurosci.* 25, 348–353
- 59 Belin, P. *et al.* (2000) Voice-selective areas in human auditory cortex. *Nature* 403, 309–312
- 60 Binder, J.R. *et al.* (1996) Function of the left planum temporale in auditory and linguistic processing. *Brain* 119, 1239–1247
- 61 Hickok, G. and Poeppel, D. (2000) Towards a functional neuroanatomy of speech perception. *Trends Cogn. Sci.* 4, 131–138
- 62 Paus, T. *et al.* (1996) Modulation of cerebral blood flow in the human auditory cortex during speech: role of motor-to-sensory discharges. *Eur. J. Neurosci.* 8, 2236–2246
- 63 Hickok, G. *et al.* (2000) A functional magnetic resonance imaging study of the role of left posterior superior temporal gyrus in speech production: implications for the explanation of conduction aphasia. *Neurosci. Lett.* 287, 156–160
- 64 Alain, C. *et al.* (2001) 'What' and 'where' in the human auditory system. *Proc. Natl. Acad. Sci. U. S. A.* 98, 12301–12306
- 65 Bushara, K.O. *et al.* (1999) Modality-specific frontal and parietal areas for auditory and visual spatial localization in humans. *Nat. Neurosci.* 2, 759–766
- 66 Warren, J.D. *et al.* (2002) Perception of sound-source motion by the human brain. *Neuron* 34, 139–148
- 67 Seltzer, B. and Pandya, D.N. (1994) Parietal, temporal, and occipital projections to cortex of the superior temporal sulcus in the rhesus monkey: a retrograde tracer study. *J. Comp. Neurol.* 343, 445–463
- 68 Bischoff-Grethe, A. *et al.* (2000) Conscious and unconscious processing of nonverbal predictability in Wernicke's area. *J. Neurosci.* 20, 1975–1981
- 69 Campbell, R. *et al.* (2001) Cortical substrates for the perception of face actions: an fMRI study of the specificity of activation for seen speech and for meaningless lower-face acts (gurning). *Brain Res. Cogn. Brain Res.* 12, 233–243
- 70 Calvert, G.A. *et al.* (2000) Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Curr. Biol.* 10, 649–657
- 71 Bogen, J.E. and Bogen, G.M. (1976) Wernicke's region – where is it? *Ann. New York Acad. Sci.* 280, 834–843
- 72 Anderson, J. *et al.* (1999) Conduction aphasia and the arcuate fasciculus: A reexamination of the Wernicke-Geschwind model. *Brain Lang.* 70, 1–12
- 73 Axer, H. *et al.* (2001) Supra- and infrasyllabic conduction aphasia. *Brain Lang.* 76, 317–331
- 74 Kuhl, P.K. (2000) A new view of language acquisition. *Proc. Natl. Acad. Sci. U. S. A.* 97, 11850–11857
- 75 Arikuni, T. *et al.* (1988) Connections of area 8 with area 6 in the brain of the macaque monkey. *J. Comp. Neurol.* 277, 21–40
- 76 Barbas, H. and Pandya, D.N. (1989) Architecture and intrinsic connections of the prefrontal cortex in the rhesus monkey. *J. Comp. Neurol.* 286, 353–375
- 77 Yeterian, E.H. and Pandya, D.N. (1998) Corticostriatal connections of the superior temporal region in rhesus monkeys. *J. Comp. Neurol.* 399, 384–402
- 78 Davis, B.L. and MacNeilage, P.F. (2000) An embodiment perspective on the acquisition of speech perception. *Phonetica* 57, 229–241
- 79 Liberman, A.M. and Whalen, D.H. (2000) On the relation of speech to language. *Trends Cogn. Sci.* 4, 187–196
- 80 Lindblom, B. (1996) Role of articulation in speech perception: clues from production. *J. Acoust. Soc. Am.* 99, 1683–1692
- 81 Rizzolatti, G. and Arbib, M.A. (1998) Language within our grasp. *Trends Neurosci.* 21, 188–194

- 82 Kluender, K.R. and Lotto, A.J. (1999) Virtues and perils of an empiricist approach to speech perception. *J. Acoust. Soc. Am.*, 503–511
- 83 Doupe, A.J. and Kuhl, P.K. (1999) Birdsong and human speech: common themes and mechanisms. *Annu. Rev. Neurosci.* 22, 567–631
- 84 Kuhl, P.K. (1994) Learning and representation in speech and language. *Curr. Opin. Neurobiol.* 4, 812–822
- 85 Gallese, V. *et al.* (1996) Action recognition in the premotor cortex. *Brain* 119, 593–609
- 86 Ferrari, P.F. *et al.* (2001) Mirror neurons for mouth actions in monkey ventral premotor cortex. *Soc. Neurosci. Abstr.* 27, 7294
- 87 Kohler, E. *et al.* (2001) Auditory mirror neurons in the ventral premotor cortex of the monkey. *Soc. Neurosci. Abstr.* 27, 1299
- 88 Watkins, K. *et al.* Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia* (in press)
- 89 Fadiga, L. *et al.* (2002) Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *Eur. J. Neurosci.* 15, 399–402
- 90 Monsell, S. (1987) On the relation between lexical input and output pathways for speech. In *Language Perception and Production: Relationships between Listening Speaking Reading and Writing* (Allport, A. *et al.*, eds), pp. 273–311, Academic Press
- 91 Darwin, C. and Hukin, R. (1998) Perceptual segregation of a harmonic from a vowel by interaural time difference in conjunction with mistuning and onset asynchrony. *J. Acoust. Soc. Am.* 103, 1080–1084
- 92 Burton, M.W. *et al.* (2000) The role of segmentation in phonological processing: an fMRI investigation. *J. Cogn. Neurosci.* 12, 679–690
- 93 Goodale, M.A. and Milner, A.D. (1992) Separate visual pathways for perception and action. *Trends Neurosci.* 15, 20–25
- 94 Petrides, M. and Pandya, D.N. (1994) Comparative architectonic analysis of the human and macaque frontal cortex. In *Handbook of Neuropsychology* (Vol. 9) (Boller, F. and Grafman, J., eds) pp. 17–58, Elsevier Science

Articles of interest in other *Trends and Current Opinion* journals

Drugs, flies, and videotape: the effects of ethanol and cocaine on *Drosophila* locomotion

Adrian Rothenfluh and Ulrike Heberlein (December 2002)

Current Opinion in Neurobiology 12, 639–645

Learning to see faces and objects

Michael J. Tarr and Yi D. Cheng (January 2003)

Trends in Cognitive Sciences 7, 23–30

Neurological disease: UPS stops delivering!

Richard J. Miller and Scott M. Wilson (January 2003)

Trends in Pharmacological Sciences 24, 18–23

Abdominal vagal afferent neurones: an important target for the treatment of gastrointestinal dysfunction

Paul L.R. Andrews and Gareth J. Sanger (December 2002)

Current Opinion in Pharmacology 2, 650–656

The pharmacology of gastrointestinal nociceptive pathways

L.A. Blackshaw and G.F. Gebhart (December 2002)

Current Opinion in Pharmacology 2, 642–649

The evolution of feeding motor patterns in vertebrates

Peter C. Wainwright (December 2002)

Current Opinion in Neurobiology 12, 691–695

Mouse models of neurofibromatosis type I: bridging the GAP

Rui M. Costa and Alcino J. Silva (January 2003)

Trends in Molecular Medicine 10.1016/S1471-4914(02)00008-4

Techniques: Applications of the nerve–bouton preparation in neuropharmacology

Norio Akaike and Andrew J. Moorhouse (January 2003)

Trends in Pharmacological Sciences 24, 44–47

Why do some perceptual illusions affect visually guided action, when others don't?

David Milner and Richard Dyde (January 2003)

Trends in Cognitive Sciences 7, 10–11