Identification of a pathway for intelligible speech in the left temporal lobe

Sophie K. Scott,[1] C. Catrin Blank,[3] Stuart Rosen[2] and Richard J. S. Wise[3]

[1]*Institute of Cognitive Neuroscience and* [2]*Department of Phonetics and Linguistics, University College London,* [3]*MRC Clinical Sciences Centre, Cyclotron Unit, Hammersmith Hospital, London, UK*

Correspondence to: Sophie K. Scott, Institute of Cognitive Neuroscience, University College London, London WC1N 3AR, UK
E-mail: sophie.scott@ucl.ac.uk

Summary

It has been proposed that the identification of sounds, including species-specific vocalizations, by primates depends on anterior projections from the primary auditory cortex, an auditory pathway analogous to the ventral route proposed for the visual identification of objects. We have identified a similar route in the human for understanding intelligible speech. Using PET imaging to identify separable neural subsystems within the human auditory cortex, we used a variety of speech and speech-like stimuli with equivalent acoustic complexity but varying intelligibility. We have demonstrated that the left superior temporal sulcus responds to the presence of phonetic information, but its anterior part only responds if the stimulus is also intelligible. This novel observation demonstrates a left anterior temporal pathway for speech comprehension.

Introduction

This study was designed to separate the left temporal lobe system responsible for the speech-specific processing of familiar, intelligible words from more general auditory processing, using functional neuroimaging (PET). A distinctly lateralized temporal lobe system for speech perception has not been demonstrated previously with functional imaging techniques, despite a wealth of clinical evidence that aphasia arises from damage to the left, not the right, hemisphere. Instead, hearing speech, contrasted with silent rest, simple tones or noise, results in activations in the left and right superior temporal cortices, sometimes associated with relative interhemispheric asymmetry (e.g. Wise *et al.*, 1991, 1999; Zatorre *et al.*, 1992; Demonet *et al.*, 1992; Binder *et al.*, 1997; Papathanassiou *et al.*, 2000). Studies in which vocalizations were contrasted with signals of similar temporal complexity also showed predominantly symmetrical activation, located in the ventrolateral superior temporal gyrus as far as the dorsal bank of the superior temporal sulcus (Mummery *et al.*, 1999; Belin *et al.*, 2000). However, interhemispheric symmetry in an increase in regional cerebral blood flow on PET images does not mean symmetry of acoustic processing functions.

In a comprehensive synthesis of the neuroimaging literature, Binder and colleagues outlined a model of speech processing in the left temporal lobe (Binder *et al.*, 1996;

Binder and Frost, 1998). They suggested that a ventrolateral stream of acoustic information from the superior to the middle and inferior temporal gyri occurs during speech perception. Importantly, at the level of the superior temporal sulcus the response was still claimed to be not speech-specific, the response at this level arising from the complex frequency and amplitude modulations that characterize speech. It was argued that speech-specific lexical and semantic processing were functions of the cortex ventral to the superior temporal sulcus.

This argument is critically dependent on the nature of the complexity of the speech and the baseline stimuli used in the experiments. Speech is an immensely complex stimulus (Pickett, 1999), from which acoustic phonetic features must be processed before they become intelligible. 'Intelligibility' refers to the comprehensibility of a signal: a fully intelligible signal could be understood and repeated by a skilled speaker of the relevant language. This term covers several properties of language, including word-form recognition, syntax and semantics. No one acoustic cue determines the intelligibility of speech, and skilled listeners are able to gather meaning from very degraded input (Miller, 1951; Shannon *et al.*, 1995). Therefore, designing stimuli that are as acoustically complex as speech but lack phonetic features and, hence, the potential for intelligibility, is difficult. In this study, we used

an established technique (Blesser, 1972) to destroy the intelligibility of two types of intelligible speech, whilst holding the structural acoustic complexity constant. Thus the non-speech stimuli were very similar to the speech stimuli in their internal structure (e.g. they contained formant-like acoustic features). This enabled us to investigate which brain regions are activated solely by intelligible speech, regardless of stimulus complexity.

Methods

Stimuli

All stimuli were based on natural sentences recorded by a single male speaker. The original unprocessed speech formed the stimuli in one condition (Sp). The other condition involving intelligible speech used what we have termed 'noise-vocoded speech' (VCo), as described by Shannon and colleagues (Shannon *et al.*, 1995). In a channel vocoder, speech is passed through a filter bank from which the time-varying envelopes associated with the energy in each spectral channel are extracted (Flanagan, 1972). Speech is reconstructed from these envelopes by periodic and aperiodic excitations that correspond to the sequence of periodicity and aperiodicity in the original speech. In noise-vocoded speech, the source of excitation is always a white noise. The result sounds like a harsh whisper, but is readily comprehensible after a brief training session. Although there are temporal fluctuations in the noise-vocoded speech waveforms that reflect vocal fold periodicity, the saliency of this pitch is quite weak (Faulkner *et al.*, 2000).

The two unintelligible conditions involved the spectral rotation (or inversion) of a signal. Rotated speech (RSp) involved spectral inversion of the original speech. This sounds like an 'alien' language: it has very similar temporal and spectral complexity to ordinary speech, but it is not intelligible. As demonstrated by Blesser (Blesser, 1972) the rotated speech (RSp) does contain some phonetic features (for example, voiceless fricatives are readily identifiable as such; voiced and unvoiced sounds are still clearly distinguishable). After extensive training, over the order of weeks, participants can learn to extract some meaning—without such training the speech is unintelligible.

Rotated speech also preserves intonation. During voiced segments, normal speech is quasi-periodic, and its spectrum can thus be approximated as a set of discrete components at multiples of the fundamental frequency (harmonics). Rotated voiced speech still has spectral components which are equally spaced in frequency, but these components are typically not multiples of some fundamental frequency—the signal is no longer truly periodic. However, such equally spaced spectral components still lead to a reasonably strong sensation of perceived pitch (Blesser, 1972), and hence its linguistic correlate, intonation.

Finally, rotated noise-vocoded speech (RVCo) is obtained by noise-vocoding of the rotated speech. This sounds like intermittent fluctuating static, with weak pitch changes; it is not at all like speech and is unintelligible.

Spectrograms of all four stimuli are shown in Fig. 1. The main acoustic difference between the speech stimuli and the noise-vocoded stimuli is the presence of quasi-periodicity (caused by vibration of the vocal folds) in the speech sounds. Perceptually, this gives both the speech and rotated speech a buzziness in which clear differences in pitch can be heard, and hence the linguistic correlate of melodic changes in speech, intonation. Thus, we also had the opportunity to investigate which cortical regions support the perception of stimuli with perceived variations in pitch, irrespective of intelligibility.

Signal processing

All stimulus materials were drawn from low-pass-filtered (3.8 kHz) digital representations of a recording of the BKB sentence lists (Foster *et al.*, 1993), binaurally presented over headphones. Spectral rotation (around 2 kHz) used a digital version of the simple modulation technique described by Blesser (Blesser, 1972). The speech signal was first equalized with a filter (essentially high-pass) that gave the rotated signal approximately the same long-term spectrum as the original. The equalized signal was then amplitude-modulated by a sinusoid at 4 kHz, followed by low-pass filtering at 3.8 kHz. Noise-vocoding was applied to each of the two signals— spectrally rotated and normal speech—using the technique described by Shannon and colleagues (Shannon *et al.*, 1995). The input waveform was passed through a bank of six analysis bandpass filters with frequency responses that crossed 3 dB down from the passband peak. Filter cut-off frequencies were obtained by dividing the frequency range from 70 to 4000 Hz equally, by the use of an equation that relates frequency to its representation on the basilar membrane (Greenwood, 1990). Envelope detection occurred at the output of each analysis filter by half-wave rectification and low-pass filtering at 320 Hz. These envelopes were then multiplied by a white noise, and each was filtered by an output filter identical to the analysis filter. The r.m.s. (root mean square) level from each output filter was then set to be equal to the r.m.s. level of the original analysis outputs, before being summed.

Pretraining

Before scanning, subjects were presented with examples of each type of stimulus. They were trained to understand the noise-vocoded speech by listening to a sentence. If they could not repeat it, they were told what the sentence was. The stimulus was then played again, until they agreed they could hear the correct sentence. This was repeated for 20 different sentences, after which each subject had comfortably reached the criterion of accurately reporting each sentence on the first presentation, without any prompting. None of the training sentences was presented in the PET study.

PET scanning

Eight right-handed normal volunteers were studied with a Siemens HR++ (966) PET scanner operated in high-
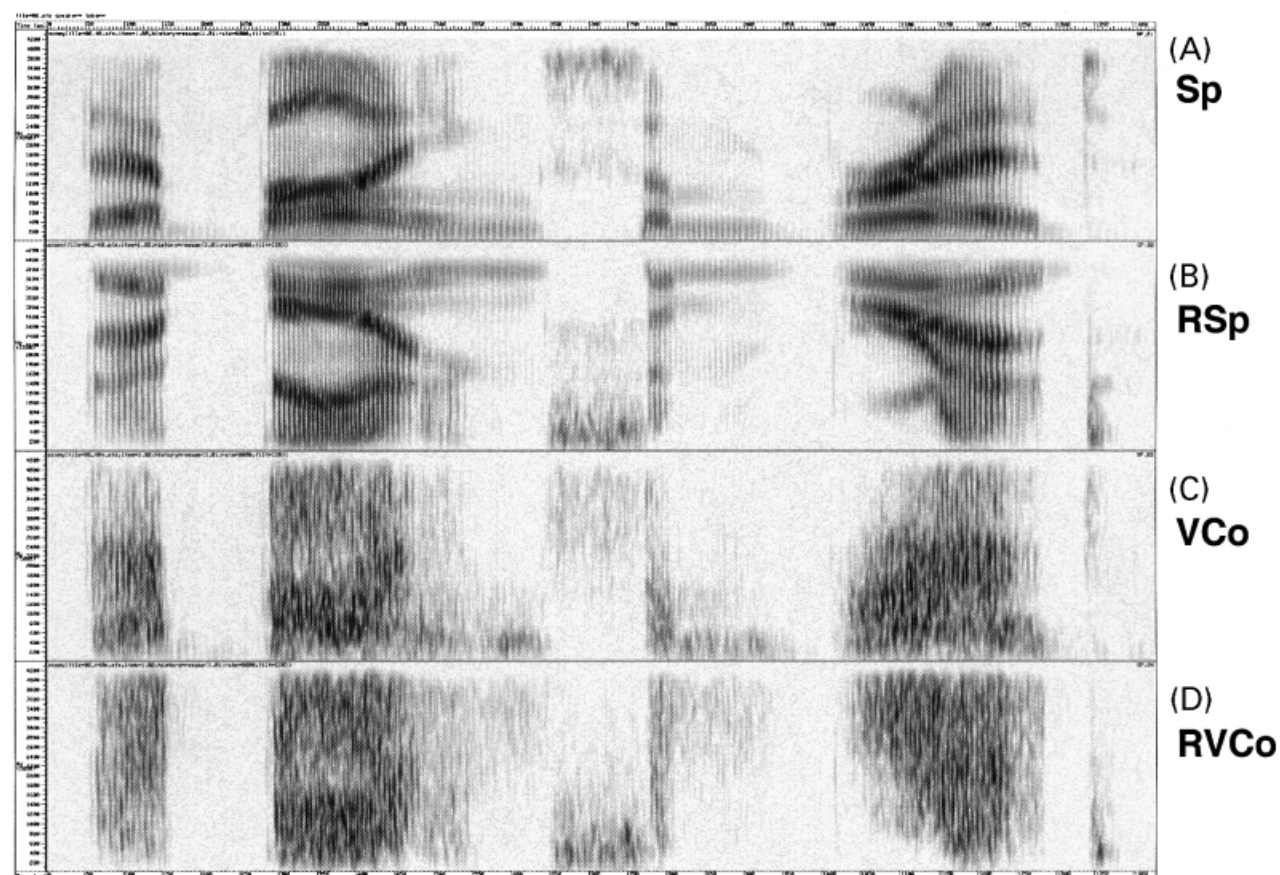
Fig. 1 Spectrograms of 'They're buying some bread'. Time is represented on the abscissa (0.0–1.43 s) and frequency on the ordinate (0.0–4.4 kHz). The darkness of the trace in each time/frequency region is controlled by the amount of energy in the signal at that particular frequency and time. (**A**) Normal speech (Sp) is intelligible with clear intonation. (**B**) Spectrally rotated speech (RSp) is not intelligible without extensive training, though some phonetic features and some of the original intonation are preserved. (**C**) Noise-vocoded speech (VCo) is intelligible, has very weak intonation and a rough sound quality. (**D**) Spectrally rotated noise-vocoded speech (RVCo) is completely unintelligible and does not sound like a voice.

sensitivity 3D mode. Each person gave informed consent before participation in the study, which was approved by the Research Ethics Committee of Imperial College School of Medicine/Hammersmith, Queen Charlotte's and Chelsea and Acton Hospitals. Permission to administer radioisotopes was given by the Department of Health (UK).

Sixteen scans were performed on each subject, using the $H_2^{15}O$ bolus technique. All subjects were scanned whilst lying supine in a darkened room with their eyes closed. There were four scans for each stimulus condition, presented in random order. Each stimulus presentation began with 20 s of varied stimuli (intelligible and unintelligible), followed by a blocked set of stimuli from just one condition that coincided with the onset of scanning. Each sentence presented was novel (i.e. there were no repeats). After each scan each subject was asked roughly how much they had understood of the stimuli they had just heard.

Analysis

The images were analysed by statistical parametric mapping (SPM99b, Wellcome Department of Cognitive Neurology;

http://www.fil.ion.ucl.ac.uk/spm), which allowed manipulation and statistical analysis of the grouped data. All scans from each subject were realigned to eliminate head movements between scans, and were normalized into a standard stereotaxic space. Images were then smoothed using an isotropic 10 mm, full-width half-maximum Gaussian kernel, to allow for variation in gyral anatomy and to improve the signal-to-noise ratio. Specific effects were investigated, voxel-by-voxel, using appropriate contrasts to create statistical parametric maps of the t statistic, which were subsequently transformed into Z scores. The analysis included a blocked ANCOVA (analysis of covariance) with global counts as confound to remove the effect of global changes in perfusion across scans. The threshold for significance was set at $P < 0.05$, corrected for analyses across the whole volume of the brain ($P < 0.000001$ uncorrected; $Z > 4.7$).

Results

All the subjects reported understanding 'all' or 'most' of the stimuli in the two intelligible conditions, and all reported making no sense of the unintelligible conditions (which were

frequently labelled 'rubbish'). The lack of intelligibility of the rotated stimuli was in accordance with previous findings (Blesser, 1972).

Two main contrasts were performed, the first to reveal regions associated with intelligibility [(Sp + VCo) – (RSp + RVCo)] and the second to reveal regions predominantly associated with a clear sensation of pitch [(Sp + RSp) – (VCo + RVCo)]. A third contrast was performed to identify areas activated by signals that contained any phonetic information, regardless of intelligibility [(Sp + VCo + RSp) – (RVCo)].

The results are presented in Fig. 2. In the left hemisphere, the superior temporal gyrus, lateral and anterior to the primary auditory cortex, and the posterior superior temporal sulcus were activated by the presence of phonetic cues and features in the signal (Sp, VCo and RSp). In contrast, the left anterior superior temporal sulcus was activated only by intelligible signals (Sp and VCo). There was a transition between these two response profiles in the mid-superior temporal sulcus. The right ventrolateral superior temporal gyrus, anterior to the primary auditory cortex, was activated by signals that had dynamic pitch variation (Sp and RSp).

The coordinates for the peaks of the activated regions were taken from the analysis software, SPM99. This coordinate system is based on the stereotaxic space devised by the MNI (Montreal Neurological Institute), itself based on the average of 305 MRI scans of normal volunteers (Evans *et al.*, 1993). It is not identical to the coordinate system of the atlas of Talairach and Tournoux (1988), although an algorithm for an approximate conversion is available (http://www.mrc-cbu.cam.ac.uk/Imaging/mnispace.html). Rather than perform this conversion and relate the coordinates of the peaks to the atlas of Talairach and Tournoux, we overlaid the activated regions on the template available in SPM99, comprising the averaged T_1-weighted MRI of 125 normal
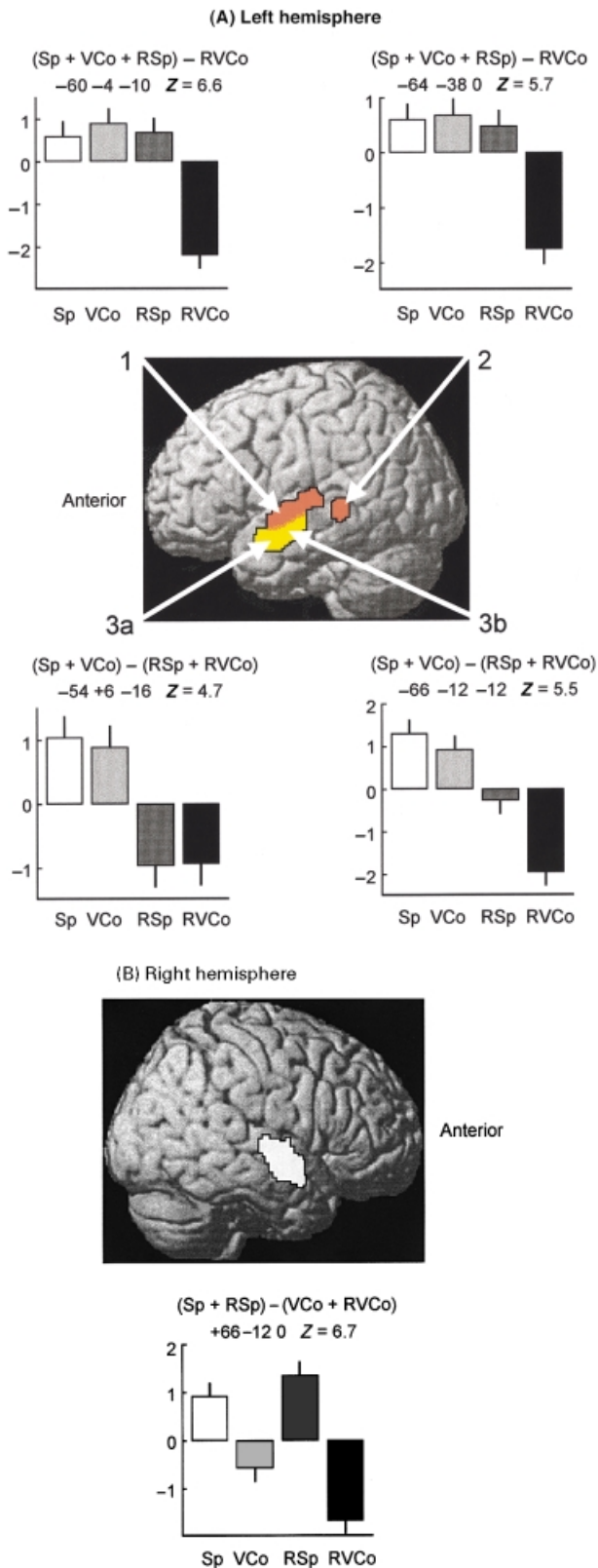

Fig. 2 Significant voxels from the three contrasts used in the analysis, co-registered on to the left (**A**) and right (**B**) lateral MRI templates that are available in the image analysis software (SPM99b). The threshold was set at $P < 0.00001$ uncorrected, excluding clusters with <50 adjacent voxels. At this threshold, the activations were confined to the temporal lobes. Each contrast was centred around zero, and the ordinate of each plot is the mean size of the effect for each condition ± standard error of the mean, within the peak voxel. The coordinates of the peak voxel (x, y and z) in the stereotaxic space of SPM99b, and the Z-score, are shown at the head of each plot. (**A**) Left temporal lobe. The contrast {(Sp + VCo + RSp) – RVCo}, colour-coded in red, showed (1) the left superior temporal gyrus, both lateral and anterior to the primary auditory cortex, and (2) a separate region in the posterior superior temporal sulcus. The contrast {(Sp + VCo) – (RSp + RVCo)}, colour-coded in yellow, showed clear separation of intelligible from non-intelligible stimuli in the anterior superior temporal sulcus (3a). In the mid-superior temporal sulcus (3b), the response profile to the stimuli appeared to be transitional between the profiles in the posterior and anterior superior temporal sulcus. This 'transition' pattern may reflect a change in response to the rotated speech stimuli, but could also arise as a result of the smoothing applied to the data. (**B**) Right temporal lobe. The only significant voxels were revealed by the contrast {(Sp + RSp) – (VCo + RVCo)}, colour-coded in white. They were located in the lateral superior temporal gyrus, anterior to the primary auditory cortex.
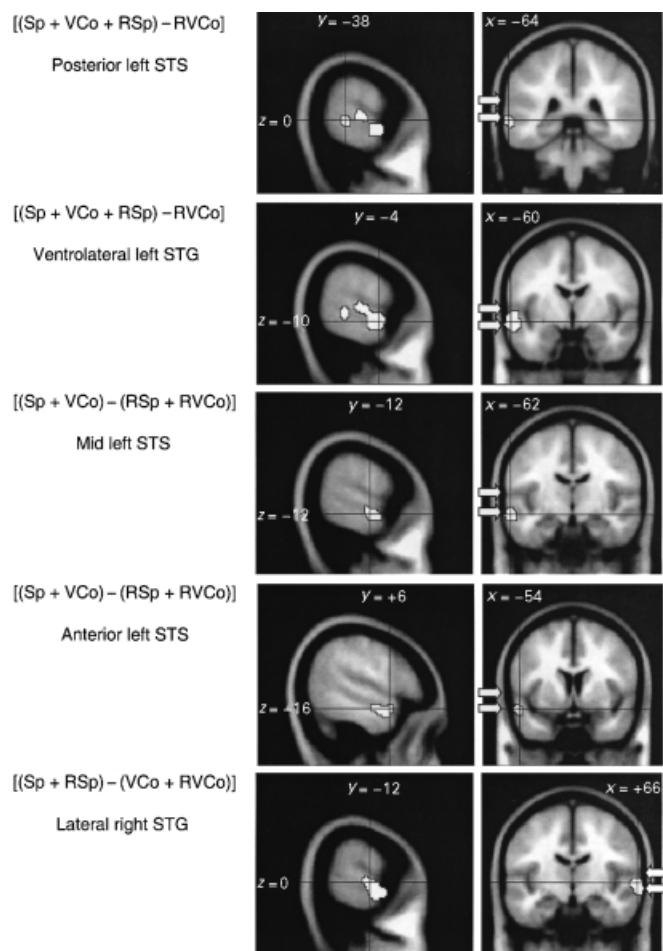
[(Sp + VCo + RSp) – RVCo]

Posterior left STS

[(Sp + VCo + RSp) – RVCo]

Ventrolateral left STG

[(Sp + VCo) – (RSp + RVCo)]

Mid left STS

[(Sp + VCo) – (RSp + RVCo)]

Anterior left STS

[(Sp + RSp) – (VCo + RVCo)]

Lateral right STG

Fig. 3 The five peak activations illustrated in Fig. 2A and B mapped on to sagittal (left column) and coronal (right column) slices of the T_1-weighted MRI template. The upper arrow points at the sylvian sulcus, the lower arrow to the superior temporal sulcus. The contrasts which reveal these activations are shown for each pair of images. The *x, y* and *z* coordinates are in millimetres relative to the plane of the anterior commissure. STS = superior temporal sulcus; STG = superior temporal gyrus.

volunteers, normalized into MNI stereotaxic space. It is evident from the sagittal and coronal planes displayed in Fig. 3 that the peaks lay in the superior temporal sulcus or, in the case of the right temporal lobe signal, the ventrolateral superior temporal gyrus. The resolution of PET and the averaging of group data mean that it was not possible to distinguish whether the superior temporal sulcus activations were located in the dorsal or the ventral bank of the sulcus.

Discussion

This study is the first clear demonstration of a left hemisphere preference for intelligible signals in a passive listening task. This lateralization is consistent with the neuropsychological literature (Caplan, 1987). The novel finding is that intelligible speech is associated with an anterolateral stream of neural information from the primary auditory cortex. The left superior temporal gyrus, ventrolateral to the primary auditory

cortex, was activated equally by speech, rotated speech and noise-vocoded speech [(Sp + VCo + RSp) – (RVCo)]. Anterior and ventral to this, the superior temporal sulcus was activated by intelligible speech only [(Sp + VCo) – (RSp + RVCo)]. In the human brain, reciprocal, monosynaptic connections from the primary auditory cortex are directed towards the lateral and anterior auditory association cortex (Galuske *et al.*, 1999), the anatomical correlate of our functional observation. In addition, our result parallels the observations from single-cell recordings in non-human primates, which show a similar stream of more complex processing, with increasing specialization of neurones in the anterolateral auditory association cortex for species-specific vocalizations (Rauschecker, 1998; Kass and Hackett, 1999; Romanski *et al.*, 1999). In humans, a similar posterior–anterior stream of neural processing of verbal representational forms, from low-level cues up to more complex, high-level constructs, is seen in the ventral temporal lobe in response to written words (Nobre *et al.*, 1994).

Several previous studies have not shown a difference between speech and baseline stimuli in the superior temporal gyrus (Binder and Frost, 1998). The complex baseline stimuli are frequently reversed speech, spoken non-words, syllables and foreign words, which all contain phonetic features and cues (such as voicing, some manners of articulation, e.g. fricatives, etc.). Non-words, syllables and foreign words can be sequenced perceptually and repeated aloud, and all can become familiar lexical–semantic items with training. If the left superior temporal gyrus is involved in the prelexical processing of phonetic cues and features and in their sequencing—processes which must occur before whole-word representation—it will be activated strongly by all these stimuli. The extent to which this region will be activated by environmental noises will depend entirely on their source. Thus, vocalized sounds (like a dog barking) will have some structural features in common with human non-speech vocalizations, and would activate this region (Morris *et al.*, 1999).

The left superior temporal sulcus, posterior to the plane of the primary auditory cortex, was also equally activated by speech, rotated speech and noise-vocoded speech, a response profile indistinguishable from that observed in the left superior temporal gyrus ventrolateral to the primary auditory cortex. The anatomical connectivity of the posterior superior temporal cortex is different from that of the anterior region, being separated from the primary auditory cortex by at least two synapses (Galuske *et al.*, 1999). This region responded to stimuli that contained some phonetic information but were not necessarily intelligible. We speculate that this area is involved in the short-term representation of sequences of sounds in a potentially pronounceable word, which is central to the ability to repeat and rehearse novel words; as a result of these processes, long-term lexical memories of familiar words are acquired (Hartley and Houghton, 1996). This speculation is based on the finding that lesions constrained to this region result in conduction aphasia, in which the

patient can comprehend speech but not repeat words (Hickok *et al.*, 2000).

For the speech and rotated speech stimuli, there was increased activity in the right ventrolateral superior temporal gyrus, anterior to the primary auditory cortex. Relative to the noise-vocoded signals, the speech signals have a strong perceived pitch and clear intonation. Our design cannot discriminate between these two levels of effect. However, the response of the right superior temporal gyrus for sounds with clear changes in pitch and/or intonation, irrespective of intelligibility, is consistent with the same right-lateralized activation seen in previous studies, in which speech or musical sequences were contrasted with noise bursts (Zatorre *et al.*, 1992, 1994) or signal-correlated noise (Mummery *et al.*, 1999; Belin *et al.*, 2000). The activity on the right is not simply a consequence of increased saliency of pitch in the speech stimuli. An elegant study of temporal pitch perception showed no clear dominance of the right temporal cortex when sounds with increasing degrees of pitch were presented (Griffiths *et al.*, 1998); instead, there was bilateral activation of the auditory cortex with increasing perceptual pitch in the signal. Thus, it seems likely that the right superior temporal gyrus preferentially processes signals with dynamic pitch variation, which is again consistent with the neurological evidence (Johnsrude *et al.*, 2000).

Therefore, this study is a clear demonstration of a speech-specific response in left temporal lobe structures. It suggests the existence of distinct subsystems within the auditory cortex, phonetic processing being principally carried out in the left superior temporal gyrus and the perception of dynamic pitch variation (in music and speech) being dependent on processes in the right superior temporal gyrus. The results demonstrate that the dorsal–ventral model of human language processing (Binder *et al.*, 1996; Binder and Frost, 1998) is correct, but this model does not address the importance of a second axis in the processing of speech—the functional transition from posterior to anterior in the left superior temporal sulcus in response to intelligible speech. We have equated the human processing of intelligible spoken language to the paralinguistic processing of vocalizations by non-human primates. The anterior superior temporal sulcus activation is strong support for the claim, based on research in non-human primates, that there is an anterior 'what' auditory pathway, which has a role in the recognition of conspecific vocalizations. This result also has implications for the evolution of human language processing.

By demonstrating distinct neural subsystems in the auditory processing of speech in the anterior–posterior axis of the temporal lobe, our results are also relevant to understanding the consequences of the location and extent of left temporal lobe strokes on the recovery of comprehension. Although the anatomical boundary of Wernicke's area has become too broad to be meaningful (Williams, 1995), most neurologists and neuropsychologists locate the core of Wernicke's area in the superior temporal cortex posterior to the plane of the primary auditory cortex (e.g. Galaburda *et al.*, 1978).

Furthermore, access to word meaning is considered to be a major function of Wernicke's area. This implies that the auditory word meaning (the verbal 'what') pathway is directed posteriorly from the primary auditory cortex. This is contrary to the general anatomical organization of the posterior cortex, with first-, second- and third-order association cortices for both the auditory and visual modalities located progressively more anteriorly (rostrally) to the primary sensory cortex (Gloor, 1997). For example, the identification of objects and faces is dependent on a pathway of visual information directed ventrally and anteriorly; the increasingly complex mental representations necessary for their unique identification are located in the anterior ventral temporal lobe, with connections via the uncinate fasciculus to prefrontal cortex (Gloor, 1997). Our results suggest that, contrary to accepted wisdom, the same general anatomical organization underlies the comprehension of speech. The anterior superior temporal sulcus projects widely to amodal high-order association cortex and medial temporal lobe structures (Jones and Powell, 1970), diffusely distributed regions within which associative knowledge about the meaning of words (i.e. semantic memory) is most likely to be represented.

References

Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B. Voice-selective areas in human auditory cortex. Nature 2000; 403: 309–12.

Binder JR, Frost JA, Hammeke TA, Rao SM, Cox RW. Function of the left planum temporale in auditory and linguistic processing. Brain 1996; 119: 1239–47.

Binder JR, Frost JA, Hammeke TA, Cox RW, Rao SM, Prieto T. Human brain language areas identified by functional magnetic resonance imaging. J Neurosci 1997; 17: 353–62.

Binder JR, Frost MS. Functional MRI studies of language processing in the brain. Neurosci News 1998; 1: 15–23.

Blesser B. Speech perception under conditions of spectral transformation: I. Phonetic characteristics. J Speech Hear Res 1972; 15: 5–41.

Caplan D. Neurolinguistics and linguistic aphasiology: an introduction. Cambridge: Cambridge University Press; 1987.

Demonet JF, Chollet F, Ramsay S, Cardebat D, Nespoulos JL, Wise R, et al. The autonomy of phonological and semantic processing in normal subjects. Brain 1992; 115: 1753–68.

Evans AC, Collins DL, Mills SR, Brown RD, Kelly RL, Peters TM. 3D statistical neuroanatomical models from 305 MRI volumes. IEEE Nucl Sci Symp Med Imag Conf Proc IEEE 1993; 108: 1877–8.

Faulkner A, Rosen S, Smith C. Effects of the salience of pitch and periodicity information on the intelligibility of 4-channel vocoded speech: implications for cochlear implants. J Acoust Soc Am. In press 2000.

Flanagan J L. Speech analysis: synthesis and perception. 2nd ed. Berlin: Springer-Verlag; 1972.

Foster JR, Summerfield AQ, Marshall DH, Palmer L, Ball V, Rosen S. Lip-reading the BKB sentence lists: corrections for list and practice effects. Br J Audiol 1993; 27: 233–46.

Galaburda A, Sanides F, Geschwind N. Human brain. Cytoarchitectonic left–right asymmetries in the temporal speech region. Arch Neurol 1978; 35: 812–7.

Galuske RAW, Schuhmann A, Schlote W, Bratzke H, Singer W. Interareal connections in the human auditory cortex. Neuroimage 1999; 9 (6 Pt 2): S994.

Gloor P. The temporal lobe and limbic system. New York: Oxford University Press; 1997.

Greenwood DD. A cochlear frequency–position function for several species—29 years later. J Acoust Soc Am 1990; 87: 2592–605.

Griffiths TD, Buchel C, Frackowiak RSJ, Patterson RD. Analysis of temporal structure in sound by the human brain. Nat Neurosci 1998; 1: 422–7.

Hartley T, Houghton G. A linguistically constrained model of short-term memory for nonwords. J Mem Lang 1996; 35: 1–31.

Hickok G, Erhard P, Kassubek J, Helms-Tillery AK, Naeve-Velguth S, Strupp JP, et al. 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 2000; 287: 156–60.

Johnsrude IS, Penhune VB, Zatorre RJ. Functional specificity in the right human auditory cortex for perceiving pitch direction. Brain 2000; 123: 155–63.

Jones EG, Powell TP. An anatomical study of converging sensory pathways within the cerebral cortex of the monkey. Brain 1970; 93: 793–820.

Kass JH, Hackett TA. 'What' and 'where' processing in auditory cortex. Nat Neurosci 1999; 2: 1045–7.

Miller GA. Language and communication. New York: McGraw Hill; 1951.

Morris JS, Scott SK, Dolan RJ. Saying it with feeling: neural responses to emotional vocalizations. Neuropsychologia 1999; 37: 1155–63.

Mummery CJ, Ashburner J, Scott SK, Wise RJ. Functional neuroimaging of speech perception in six normal and two aphasic subjects. J Acoust Soc Am 1999; 106: 449–57.

Nobre AC, Allison T, McCarthy G. Word recognition in the human inferior temporal lobe. Nature 1994; 372: 260–3.

Papathanassiou D, Etard O, Mellet E, Zago L, Mazoyer B, Tzourio-Mazoyer N. A common language network for comprehension and production: a contribution to the definition of language epicentres with PET. Neuroimage 2000; 11: 347–57.

Pickett, JM. The acoustics of speech communication: fundamentals, speech perception theory, and technology. Boston: Allyn and Bacon; 1999.

Rauschecker JP. Cortical processing of complex sounds. [Review]. Curr Opin Neurobiol 1998; 8: 516–21.

Romanski LM, Tian B, Fritz J, Mishkin M, Goldman-Rakic PS, Rauschecker JP. Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. Nat Neurosci 1999; 2: 1131–6.

Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M. Speech recognition with primarily temporal cues. Science 1995; 270: 303–4.

Talairach J, Tournoux P. Co-planar stereotaxic atlas of the human brain. Stuttgart: Thieme; 1988.

Williams PL. Gray's anatomy: the anatomical basis of medicine and surgery. 38th ed. New York: Churchill Livingstone; 1995.

Wise R, Chollet F, Hadar U, Friston K, Hoffner E, Frackowiak R. Distribution of cortical neural networks involved in word comprehension and word retrieval. Brain 1991; 114: 1803–17.

Wise RJ, Greene J, Büchel C, Scott SK. Brain regions involved in articulation. Lancet 1999; 353: 1057–61.

Zatorre RJ, Evans AC, Meyer E, Gjedde A. Lateralization of phonetic and pitch discrimination in speech processing. Science 1992; 256: 846–9.

Zatorre RJ, Evans AC, Meyer E. Neural mechanisms underlying melodic perception and memory for pitch. J Neurosci 1994; 14: 1908–19.