Defining a Left-lateralized Response Specific to Intelligible Speech Using fMRI

C. Narain[1,2], Sophie K. Scott[3], Richard J.S. Wise[4], Stuart Rosen[5], Alexander Leff[4], S.D. Iversen[2] and P.M. Matthews[1]

[1]Centre for the Functional Magnetic Resonance Imaging of the Brain (FMRIB), [2]Department of Experimental Psychology, University of Oxford, [3]Department of Psychology, University College London, [4]Clinical Sciences Centre, Imperial College, London, [5]Department of Phonetics and Linguistics, University College London, UK

Functional imaging studies of language have shown bilateral superior temporal activations in response to 'passive' perception of speech when the baseline condition did not control for the acoustic complexity of speech. Controlling for this complexity demonstrates speech-specific processing lateralized to the left temporal lobe, and our recent positron emission tomography study has emphasized a role for left anterolateral temporal cortex in speech comprehension. This contrasts with the more usual view that relates speech comprehension to left temporal-parietal cortex, the ill-defined area of Wernicke. This study attempted to reconcile these differences, using a more sensitive 3 T functional magnetic resonance imaging system, and a sparse sampling paradigm. We found left lateralized activations for intelligible speech with two distinct foci, one in the anterior superior temporal sulcus and the other on the posterior temporal lobe. Therefore, the results demonstrate that there are neural responses to intelligible speech along the length of the left lateral temporal neocortex, although the precise processing roles of the anterior and posterior regions cannot be determined from this study.

Keywords: acoustics, language, temporal lobe

Introduction

Historically, the posterior left temporal gyrus (STG) has been identified as crucial for understanding speech (Wernicke, 1874). From a psychoacoustic perspective, mapping the spectral and temporal properties of the speech signal on to long term mental representations of meaning involves a range of processes which may be organized in series, in parallel or both (Scott and Johnsrude, 2003; Scott and Wise, 2003b) and which may occur prior to later semantic and syntactic analysis. The study of the acoustic basis of speech perception is commonly referred to as speech intelligibility or speech reception (Miller, 1951). Studies using this methodology typically use the number of key words repeated from heard simple sentences as an index of the degree of accuracy with which the acoustic speech signal has been decoded (Blamey *et al.*, 2001; Keidser and Grant, 2001). Thus speech intelligibility has been used to characterize the effects of numbers of channels in noise-vocoded speech (Shannon *et al.*, 1995), the effects of speech presented in different noise contexts (Brungart *et al.*, 2001; van Wijngaarden *et al.*, 2002), and the impact of age-related and pathological hearing loss (Peters *et al.*, 1998). This term thus describes a signal in which there is sufficient acoustic detail for a listener to decode whole words. It makes no assumptions about the specific phonetic information needed to do this, not least because there is no simple linear mapping between acoustic cues and phonetic identity (Bailey and Summerfield, 1980). The term intelligibility is not aimed at dissecting the relative contributions of semantic and syntactic information to the decoding of the speech signal and is construed within an acoustic rather than linguistic framework.

Speech perception is also intimately linked with speech production, and the ability to repeat is to some degree dissociable from comprehension. Thus non-words can be repeated but are not understood. There is evidence that the implicit rehearsal of heard speech occurs during speech perception, generally conceived as a sound-to-articulation pathway that operates in parallel with a sound-to-meaning pathway (Hickok and Poeppel, 2000). This has also been described in neural models of speech perception (Hickok and Poeppel, 2000; Wise *et al.*, 2001; Scott and Johnsrude, 2003; Scott and Wise, 2003b).

Functional imaging of passive speech perception (where no overt response is required) will demonstrate all the implicit processes associated with comprehension and rehearsal (repetition), but the power of any particular study may emphasize one route of speech processing and not the other. This is particularly so when statistical thresholding of the images excludes physiologically relevant signal, or the methodology used to obtain the images is not equally sensitive across the whole region of the brain under study (Devlin *et al.*, 2000). Functional magnetic resonance imaging (fMRI) provides a non-invasive method of studying language processing, with the potential for greater sensitivity than positron emission tomography (PET) (due to more possible scans per condition) but with the potential loss of signal in anterior temporal lobe structures (Devlin *et al.*, 2000).

Previous functional imaging studies of the acoustic basis of speech perception have typically shown bilateral responses to speech in the temporal lobes (Wise *et al.*, 1991; Zatorre *et al.*, 1992; Mummery *et al.*, 1999; Benson *et al.*, 2001; Vouloumanos *et al.*, 2001; Wong *et al.*, 2002) for passive listening. This is in contrast to the indication, from the clinical literature, that left temporal lobe lesions result in sensory aphasic problems (Wernicke, 1874; Turner *et al.*, 1996; Kuest and Karbe, 2002). One possible reason for this discrepancy may be the choice of the control condition, which often have not appropriately accounted for the acoustic complexity of speech. When addressing the acoustic processing of speech, this is a crucial aspect of the study design, since speech is an immensely complex acoustic signal. Speech contains quasi-periodic and aperiodic sections (due to the presence or absence of voicing), amplitude modulation, frequency modulation and considerable spectral structure (formants) due to the movements of the articulators, as well as periods of silence. Previous studies have therefore frequently shown bilateral activation that is probably due to a lack of control for the acoustic structure in the baseline condition. No one cue determines the intelligibility of speech (Miller, 1951), but a certain

Cerebral Cortex December 2003;13:1362–1368; DOI: 10.1093/cercor/bhg083

degree of spectral-temporal modulation is essential (Drullman, 1995; Shannon *et al.*, 1995; Faulkner *et al.*, 2001). In order to delineate areas only involved in processing the correlates of speech intelligibility, separately from more general acoustic processing, the control stimulus should ideally have all the acoustic properties of speech, without being intelligible. However, it is difficult to design a stimulus as acoustically complex as speech which is not intelligible (Blesser, 1972). Skilled listeners are able to understand speech with very degraded spectral and temporal input (Drullman, 1995; Shannon *et al.*, 1995).

A second limitation of many previous imaging studies is that they have used a simple subtraction paradigm, assuming that differences between two cognitive states have been well controlled for all but the variable of interest. However, this assumption may not be valid (Sartori and Umilta, 2000). An alternative approach, which gets around many of the problems of the subtraction method, is to use more than one contrast isolating the same function, and to identify overlapping regions of activation as those important for a common processing function (Price and Friston, 1997).

Two recent publications have used evidence from PET to explicitly emphasize two streams of speech processing in the left temporal lobe: a route directed towards anterolateral temporal cortex associated with the intelligibility of the stimulus (Scott *et al.*, 2000) and another directed posteriorly that, it was proposed, was associated with repetition (Wise *et al.*, 2001). The present study has used fMRI to investigate further the anterior–posterior extent of activation of left lateral temporal neocortex to speech that can be both implicitly understood and temporally sequenced for repetition. We used a technique called sparse sampling that limited the influence of scanner noise on the physiological response (Hall *et al.*, 1999). We also used two forms of intelligible speech: normal speech and noise-vocoded speech (Shannon *et al.*, 1995), and baseline stimuli that controlled for spectral and temporal structure in the acoustic signals. These two sets of stimuli were contrasted against each other using a conjunction analysis, in order to avoid the shortcomings of a simple cognitive subtraction paradigm.

Materials and Methods

Subjects

The study was conducted in accordance with the guidelines of the Central Oxford Regional Ethics Committee, and written consent was obtained from all subjects in accordance with the declaration of Helsinki. Data was collected from 11 right-handed subjects (9 male, 2 female), all of whom had English as their first language. The mean age was 27 years (age range 20–50).

Stimuli

Four different stimuli were used: normal speech (Sp), six-channel noise-vocoded speech (VCo), spectrally rotated normal speech (RSp) and spectrally rotated noise-vocoded speech (RVCo). In each case, the stimuli were sentences of ~2 s duration, which had been appropriately transformed. Sentences were taken from the Bamford–Kowal–Bench (BKB) standard sentence list (Bench *et al.*, 1979), and were simple, unconnected statements (e.g. 'The clown has a funny face'), with imageable, concrete words and very simple syntax. Short sentences like these are commonly used to determine speech intelligibility thresholds clinically and experimentally, and this also has the advantage that more signal is present than in word lists, increasing the power of the study. The use of short sentences also enhances the intelligibility of the noise-vocoded speech. All stimuli were presented using MR-compatible electrostatic headphones built by the Institute of Hearing Research in Nottingham (http://www.ihr.mrc.ac.uk/) and designed specifically for use in an MRI system.

Noise-vocoded speech breaks the speech signal down into amplitude modulation at difference frequency bandwidths. The more frequency channels that are used, the more intelligible the speech is (Shannon *et al.*, 1995), and below eight channels the relationship between number of channels and intelligibility is logarithmic (Faulkner *et al.*, 2001). With training, subjects can easily learn to understand speech with only four channels. This is because the speech has not been distorted, instead the original signal is being presented with reduced amount of information, and the subjects quickly learn what sort of information that has been preserved.

The noise-vocoded stimuli (VCo and RVCo) thus differ from the speech stimuli in three ways. First, spectral variation in these stimuli is conveyed by band passed filtered noise, rather than the quasi periodic vibrations of the human vocal folds, and therefore subjectively these stimuli sound like a harsh whisper. Second, the temporal and spectral profile of noise-vocoded speech is smeared, so that these stimuli can be said to be acoustically less complex than natural speech. However, VCo stimuli are readily intelligible after a short training session (on the order of 15 min). Third, the sense of pitch is very reduced with six-channel noise-vocoded speech, so the intonation of these stimuli is very attenuated.

The rotated speech (RSp) stimuli can be thought of as mirror images of untransformed stimuli, as high and low frequencies are inverted around a single chosen frequency (here, 2 kHz). These stimuli are distinct from reversed speech commonly used as unintelligible stimuli in functional imaging studies of speech perception. For the details of the transformations involved see Scott *et al.* (2000). Unlike reversed speech, the temporal and spectral structure has been largely preserved. Reversed speech differs from normal speech in a number of ways, overall intelligibility being just one of them. For example, whereas much of normal speech has fast onsets (e.g. plosives) and long decays, reversed speech has slow onsets and rapid decays. The phonotactic structure is affected, and sequences are generated that could not be articulated. Thus, a comparison between speech and reversed speech would pick up differences related to processing the temporal structure as well as intelligibility.

RSp stimuli, on the other hand, contain phonetic features (Blesser, 1972): thus manner of articulation is often preserved (e.g. frication is identifiable although the fricatives themselves are changed, and silence before obstruents is unaltered). All the original acoustic information is still available, though it is now in the wrong frequency channels, for example the low frequency amplitude modulations are at higher frequencies. Blesser (1972) described spectrally rotated speech as sounding like an alien speaking your language with a completely different set of articulators (for examples of stimuli, see http://www.phon.ucl.ac.uk/home/brain/). Rotated speech thus has the potential to become intelligible, although this requires extensive training (on the order of weeks and months) (Blesser, 1972; Rosen *et al.*, 2002). The RVCo stimuli sound like intermittent static, and are not at all speech-like, although they contain the same amount of acoustic information as the noise-vocoded speech. To date, there is no evidence that such stimuli can ever be understood.

Subjects in this study were pretrained on the six-channel noise-vocoded speech, to a level where they were performing at ceiling on their repetition of noise-vocoded sentences (see procedure). Therefore both the Sp and VCo stimuli are both fully intelligible (i.e. they could be understood and repeated) and the two sets of rotated stimuli are fully unintelligible (they could not be either understood or repeated) (Rosen *et al.*, 2002). The Sp stimuli are readily comprehensible on first presentation, while the VCo stimuli are comprehensible after a brief training session.

Experimental Design

The experiment consisted of ten blocks of each of the four stimulus types. Each block consisted of presentations of five consecutive sentences, each of 2 s duration, giving a total block length of 10 s. Blocks of sentences were used in preference to single sentences, as pilot work indicated that blocks of sentences of the same type produce

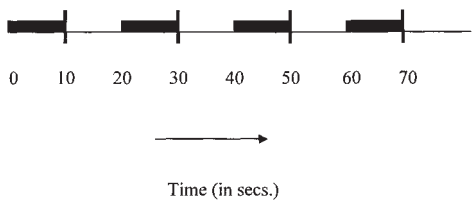

Figure 1. Schematic representation of the experiment. The shaded horizontal rectangle represents 10 s long blocks of stimuli, and the vertical line a single multislice volume. Each block began 10 s after the end of the preceding one, and the end of each block was followed by the acquisition of a multislice volume. The experimented consisted of 40 such blocks, with ten blocks of each of the four stimulus types randsomised so that no two consecutive blocks were of the same stimulus type.

a greater BOLD signal, which reaches a maximum at ~10 s after the start of the stimulus (unpublished data). Block presentation was randomized, so that no two consecutive blocks contained the same stimulus type.

Using a 'sparse' sampling MRI acquisition design (Hall *et al.*, 1999), a multislice volume covering the temporal lobes was collected at the end of each block. No scans were collected during the presentation of the stimuli themselves, so that at no time did the stimuli overlap with the scanner noise (see Fig. 1). Each block was followed by a silent period of 10 s to allow recovery of the hemodynamic response (Hall *et al.*, 1999). As the HRF was sampled at its peak, there was sufficient signal-to-noise even with the reduced number of HRF acquisitions per stimulus type.

Training

Subjects were pretrained on VCo stimuli prior to scanning. This was performed in an interactive manner: subjects were played a single VCo sentence and asked to repeat it. If subjects could not understand the sentence, or gave an incorrect or partial response (an incorrect response was defined as any word being repeated incorrectly), the sentence was played again. The subject was then asked to repeat the sentence, and if the response was still not correct, the experimenter repeated the sentence in normal speech, and played the VCo speech sentence again, at which point all subjects were able to repeat the sentence correctly. This procedure was repeated until subjects were able to understand VCo sentences clearly on first presentation, for 15 successive sentences. This ensured that the subjects were fully able to understand the noise-vocoded speech. Unlike sinewave speech (Remez *et al.*, 1981), subjects trained on VCo speech cannot reverse their perception and hear the signal as noise (Shannon *et al.*, 1995). This measure of intelligibility is thus identical to research mentioned in the introduction, and avoids confounds introduced when subjects are required to transcribe sequences (not using phonetic transcription). This method has been used to determine the 'phonetic' information in reversed speech (Binder *et al.*, 2000) but suffers from the problem that the perception and written expression of a sequence, even with phonetic transcription, is not a unitary cognitive process and distortions are introduced, not least the regularization of sounds into expressible symbols (Scott and Wise, 2003a). Using accurate repetition as an index of intelligibility has the benefit of being ecologically valid, though it may not be sensitive to other factors influencing speech perception, e.g. the newly acquired skill of decoding noise-vocoded speech might be more easily disrupted by a concurrent task. Examples of Sp, RSp and RVCo speech also were presented, to ensure that they were (i) familiar with the stimuli, and (ii) unable to understand the rotated speech and the rotated noise-vocoded speech. This training session took ~20 min, and none of the material used during training was repeated during the experiment.

Before the start of the experiment, subjects were told that they were to simply lie in the scanner, listen to the auditory stimuli that would be presented to them, and try and understand their meaning (Mummery *et al.*, 1999; Scott *et al.*, 2000). Subjects were told that it was important to pay attention to all that they heard, but they should make no explicit attempt to try and remember any of the sentences.

fMRI Scanning

The study was performed on a Varian Innova 3 T MRI/MRS system with a purpose-built birdcage radiofrequency (RF) coil. Functional scans were collected using a echoplanar imaging (EPI) sequence. A 'sparse' sampling design (Hall *et al.*, 1999) was employed, with a single multislice volume acquired over 3 s, every 20 s. Acquisition began immediately after the offset of each auditory stimulus presentation (see Fig. 1). Each volume consisted of 21 axial slices, including the entire temporal lobes, with a notional single voxel resolution of $4 \times 4 \times 5$ mm.

Data Analysis

Data analysis was carried out using statistical parametric mapping (SPM 99, Wellcome Department of Cognitive Neurology, London, UK), implemented in Matlab (Mathworks Inc. Sherborn, MA, USA). Data was regrouped so that all scans belonging to one stimulus type were treated as a single epoch. All volumes were realigned to the first collected volume, and then resliced using a sinc interpolation. They were then smoothed using a Gaussian filter of 6 mm, normalized to the MNI template in SPM99, and a group analysis was carried out using a fixed-effects model. We used the conjunction analysis option in SPM to show areas activated in common when two contrasts were designed to show the same underlying process. This is a statistically conservative test, and has been shown to have several advantages over a simple cognitive subtraction approach (Price and Friston, 1997).

Condition and subject effects were estimated using the general linear model. To test hypotheses about regionally specific effects of different stimuli, these estimates were compared using linear contrasts. The resulting set of voxel values for each contrast is an SPM of the *t*-statistic. Effects are reported as significant above a threshold of $P < 0.05$ corrected. As the results were assessed using a statistically conservative conjunction analysis, type I errors were avoided.

Results

After the experiment, the subjects were asked how many types of stimuli they had heard, and of these, which ones they understood. Finally, they were asked to rank the stimulus types according to how intelligible they were perceived to be. This was an informal confirmation of the intelligibility seen in the pretraining session. All (11) subjects reported that there were four different types of stimuli, and that they could understand two of the different types (typically referred to as 'the normal speech' and 'the one we trained with'). They rated the Sp and VCo conditions as equally intelligible, and the RSp and RVco conditions as equally unintelligible. This is consistent with the pretraining and with the behavioural literature (Blesser, 1972).

Two linear contrasts were performed to determine the neural correlates of intelligible speech, while controlling for stimulus complexity. Thus Sp was contrasted with RSp and VCo was contrasted with RVCo. In both of these contrasts, an intelligible stimulus was contrasted with an unintelligible stimulus of equal acoustic complexity. A conjunction analysis was then carried out using the conjunction option in SPM, and only voxels that were significant at a threshold of $P < 0.05$ corrected for the conjunction contrast were accepted as regions responding specifically to intelligible speech across both speech and Vco conditions.

Three significant clusters (>10 voxels) were observed, all in the left temporal lobe. Areas activated included the dorsal posterior margin of the left temporal lobe (Wernicke's area) (Bogen and Bogen, 1976; Wernicke, 1874). Anteriorly, we also found activation centred on the mid and anterior superior temporal sulcus. There was no significant activation on the

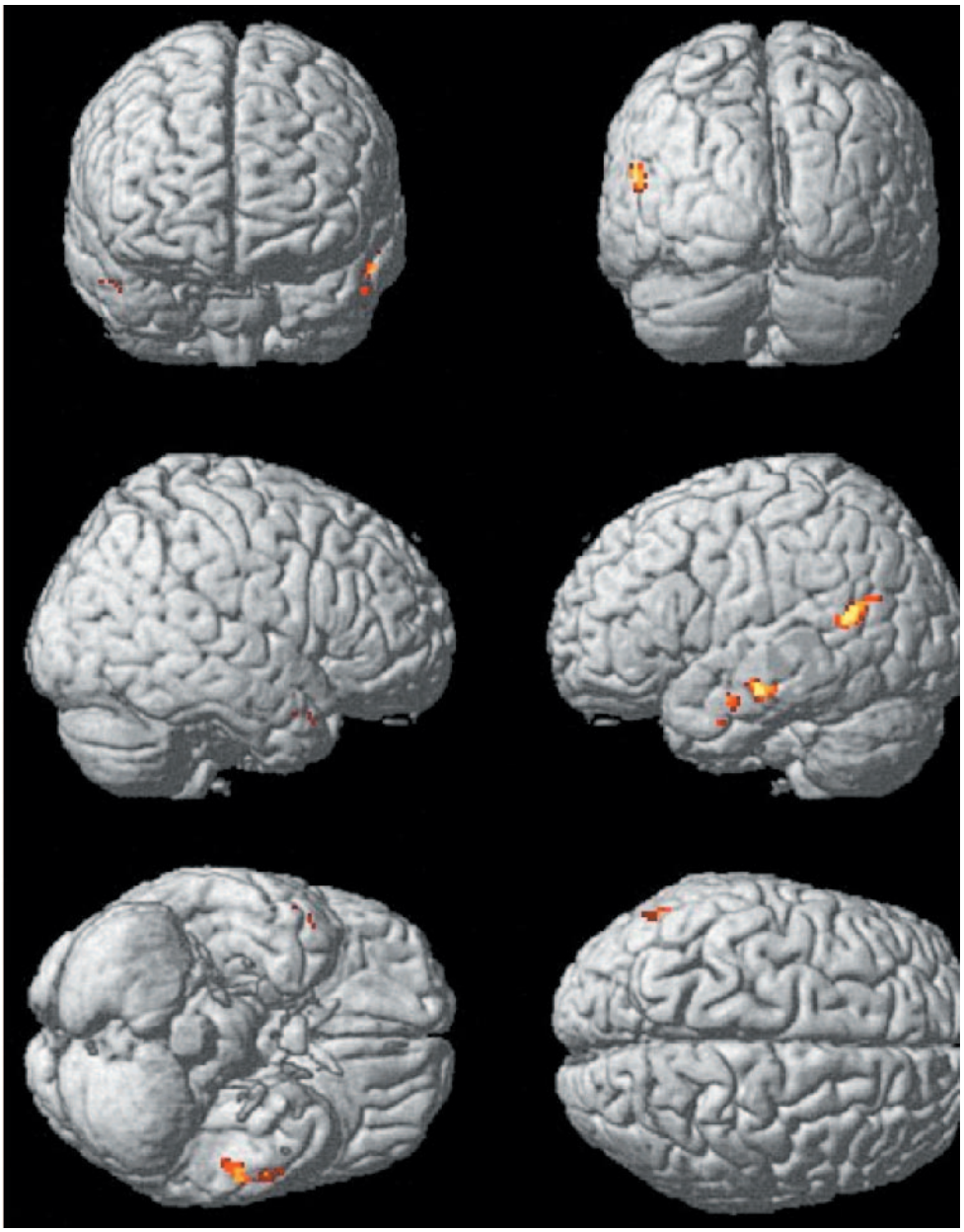

Figure 2. Maximum intensity projections of conjunction analysis between (Sp – RSp) and (Vco – RVCo). The areas in common are processing intelligibility. The figure shows the results of a fixed effects group analysis t test plotted onto an MNI template. $P = 0.05$ (corrected).

right hemisphere at the thresholds we used (0.05 corrected) (see Fig. 2 and Table 1).

Discussion

Our study was aimed at a reconciling our previous PET study, which indicated that the anterior STS was the most significant area involved in processing intelligibility, with the large body of research which assigns this role to the posterior temporal-parietal area. Most previous PET and fMRI studies of language have shown bilateral activation with passive listening tasks, despite clinical data that clearly shows that the left but not the right hemisphere is crucial in processing language. While brain imaging could identify regions modulated but not critical to speech processing, other explanations may be more tenable. Our earlier report (Scott *et al.*, 2000) provided direct evidence supporting the hypothesis that a principal reason for the relative lateralization discrepancy was the failure to use baseline conditions that adequately controlled for the complex physical properties of speech in contrasts aimed at defining processing of intelligibility. While this study demonstrated clear left lateralization for processing intelligible speech, the activity was seen in the anterior STS, not in the posterior regions emphasized by clinical lesion studies (Turner *et al.*, 1996). This result is consistent with the changes seen in semantic dementia (characterized by a progressive deterioration in the comprehension of words), which is associated with grey matter loss in the left temporal lobe (Chan *et al.*, 2001) and previous functional imaging studies which have impli-

Table 1
Results of contrasts carried out

	Contrast	*x*	*y*	*z*	*t*-statistic	No. of voxels	Anatomical area
	Conjunction (see text)						
1	(Sp+ Vco) – (RSp – RVCo)	–54	–18	–18	4.20	49	Mid STS
		–52	–54	14	4.12	81	Wernicke's area
		–56	–6	–20	3.76	13	Anterior STS

The *x*, *y* and *z* coordinates are in talaraich space, and refer to the peak voxel activated in each contrast. Peak voxels are reported for all major clusters (>10 voxels). All contrasts are thresholded at $P = 0.05$ (corrected). STS, superior temporal sulcus.

cated left anterior temporal lobe regions in the processing of intelligible speech. Thus Mazoyer *et al.* (1993) and Schlosser *et al.* (1998) showed anterior temporal lobe involvement in the perception of connected speech. The baseline conditions used were foreign languages, leaving open the possibility that these results could be driven by acoustic (including phonotactic) differences between the intelligible and unintelligible speech conditions.

However, the smaller number of scans used (four per condition) in the Scott *et al.* (2000) PET study means that this study may have missed subthreshold activation in the posterior regions which are known to have a role in speech comprehension, and which have been identified as speech specific regions of asymmetry in previous studies (Mummery *et al.*, 1999). Our current study therefore attempted to exploit the complementary sensitivity (Devlin *et al.*, 2000) and higher spatial resolution of fMRI to further explore this problem.

We confirmed the results from our PET study (Scott *et al.*, 2000) showing a strongly left-lateralized activation for intelligibility in a passive language-listening task. We also found a region within the posterior superior temporal lobe (Wernicke's area) (Bogen and Bogen, 1976) responding to intelligible speech. This study thus reconciles imaging data with the large body of evidence that aphasia arises most typically from left but not right hemisphere damage, and that Wernicke's area has a crucial role in language comprehension.

We found two distinct regions in the left temporal lobe showing greater activity associated with intelligible speech relative to a complex acoustic stimulus with speech-like characteristics: posteriorly on the superior temporal gyrus, and anteriorly on the STS. Previous studies (Zahn *et al.*, 2000) have also found activation in the STG and STS to be specific to meaningful speech, but only when subjects were required to carry out a higher order conceptual semantic task. These activations were also much more anterior to those seen in the current study. In a recent study, Friederici *et al.* (2003) demonstrated responses to semantic and syntactic violations in the left anterior and posterior STG. However, direct contrasts of the two violation conditions showed no significant differences in these regions.

Binder *et al.*'s (2000) very comprehensive study aimed to isolate brain regions activating specifically to intelligibility from those involved in lower level sound processing. Our study is comparable in terms of aims and methodology, but we found a more left lateralized network. We attribute these differences to several factors. An important reason for the difference in results is the control stimuli used. Reversed speech, as used by Binder *et al.* (2000), represents a substantial improvement over the use of simple stimuli such as tones, but this may still not be an ideal control condition. Reversed speech differs from normal speech on a number of acoustic features other than intelligibility: it has comparatively slower onsets and more rapid acoustic decays, leading to a distortion of the temporal code, which can be hard to characterize. We suggest that using a baseline condition which does not adequately control for these acoustic features may result in increased right hemisphere activation. Spectrally rotated speech is an improved control stimulus, as it still retains the temporal code and spectral complexity inherent in normal speech, but is at the same time meaningless to an untrained listener (in the absence of extensive training on the order of weeks) (Rosen *et al.*, 2002). The use of sentences rather than words also contributes to the power of the present study. Further, we present a conjunction of two different forms of speech, a statistically conservative procedure which also increases the power to detect significant activations. These factors together explain why unlike Binder *et al.* (2000) study, we find clear differences in the processing of intelligible versus unintelligible speech above the dorsal STS, a difference not uncovered by their comparison of speech and reversed speech.

Finally, auditory studies using fMRI suffer from the constant background noise due to scanner gradient switching. The potential problems due to this background noise are acknowledged by Binder *et al.* (2000), who worked to reduce this confound by minimizing slice coverage and using a relatively long interscan interval. We used an alternative data acquisition strategy called 'sparse sampling' (Hall *et al.*, 1999, which integrates these features, and at the same time ensures that the stimuli are presented in silence and that the obtained activations are not affected due to activation caused by the scanner noise. Even though the number of multi-slice volumes collected per unit time is reduced using such a paradigm, we believe that the advantages of this method may make it more suitable for auditory experiments using fMRI.

Our results showing clear left lateralization for processing intelligibility extend the Scott *et al.* (2000) PET study, on which this study was based. However, our results differ from the earlier work in that the posterior activation is stronger than the anterior (STS) activation, in direct contrast to the PET study. This is potentially due to greater sensitivity in the current study: PET is limited as to the number of volumes acquired, and thus subthreshold activations may be lost. fMRI, in contrast, affords many more data points to be collected for each condition (even with sparse sampling), which can improve the power of analysis. To test this hypothesis, we reanalysed our previous PET data and found that there was subthreshold activation [$z = 4.33$, P(corrected) = 0.29] in the posterior STG (MNI coordinates: –52 –58 8) (S.K. Scott and R.J.S. Wise, unpublished data).

The argument in the Scott *et al.* (2000) study was based around the strong activation seen in the anterior STS, especially since this ran forward from other STG regions that were activated by stimuli with the acoustic correlates of phonetic cues and features. This result is seen only weakly in the present study, and a number of other studies using fMRI have tended to emphasize a temporoparietal (Wernicke's) area

(Binder *et al.*, 1997, 2000; Calvert *et al.*, 1997; Zahn *et al.*, 2000) as being important in semantic processing. However, the results of fMRI studies of language may be influenced by the fact that the anterior STS is relatively more susceptible to dramatic signal loss due to magnetic susceptibility artifacts (Devlin *et al.*, 2000). The anterior STS activation described in this paper may have been affected by such artefacts.

In terms of functional significance, a reanalysis of several PET studies (Wise *et al.*, 2001) has identified activity in posterior STS associated with a variety of language based tasks, not solely speech comprehension. A PET study (Crinion *et al.*, 2003) suggests that Wernicke's area activates more significantly in an intelligibility task related to processing stories rather than isolated sentences. An ERP study by Abdullaev and Posner (1998) found that activation in the Wernicke's area occurred later than the behavioural response in a semantic decision task. Therefore, we suggest that temporal-parietal junction may form the part of a short-term memory network specialized for language. We could therefore expect it to be more activated with processing connected sentences (as in stories) than by single sentences, and more by sentences than by single words. This would also explain its relatively late activation after stimulus presentation, and its reduced involvement in studies using single word stimuli (though see Mummery *et al.*, 1999). Our hypothesis is supported by suggestions that language may have evolved from such a working memory network, and that Wernicke's area (posterior STG) may be the focus of a multi-modal network associated with language comprehension (Aboitiz and Garcia, 1997).

Based on the converging evidence from our previous PET study and the present study, we therefore suggest that the areas we have uncovered (the anterior STS and the posterior STG) are part of a distributed system of regions associated with the comprehension of speech. Further, the anterior STS and Wernicke's area may be serving different functions in such a system.

Notes

This work was conducted at the Oxford Centre for the Functional Magnetic Resonance Imaging of the Brain (FMRIB). P.M.M. thanks the MRC for personal support and support for the Centre. R.J.W., S.K.S. and C.N. gratefully acknowledge support from the Wellcome Trust. C.N. would also like to thank all the subjects who took part in the study.

Address correspondence to Charvy Narain, FMRIB Centre, John Radcliffe Hospital, Headington, Oxford OX3 9DU, UK. Email: charvy@fmrib.ox.ac.uk.

References

Abdullaev YG, Posner MI (1998) Event-related brain potential imaging of semantic encoding during processing single words. Neuroimage 7:1–13.

Aboitiz F, Garcia V (1997) The evolutionary origin of the language areas in the human brain. A neuroanatomical perspective. Brain Rev 25:381–396.

Bailey PJ Summerfield Q (1980) Information in speech: observations in perception. J Exp Psychol Hum Percept Perform 6:536–563.

Bench J, Kowal A, Bamford J (1979) The BKB (Bamford–Kowal–Bench) sentence lists for partially-hearing children. Br J Audiol 13:108–112.

Benson RR, Whalen DH, Richardson M, Swainson B, Clark VP, Lai S, Liberman AM (2001) Parametrically dissociating speech and nonspeech perception in the brain using fMRI. Brain Lang 78:364–396.

Binder JR, Frost JA, Hammeke TA, Cox RW, Rao SM, Prieto T (1997) Human brain language areas identified by functional magnetic resonance imaging. J Neurosci 17:353–362.

Binder JR, Frost JA, Hammeke TA, Bellgowan PS, Springer JA, Kaufman JN, Possing ET (2000) Human temporal lobe activation by speech and nonspeech sounds. Cereb Cortex 10:512–528.

Blamey PJ, Sarant JZ, Paatsch LE, Barry JG, Bow CP, Wales RJ, Wright M, Psarros C, Rattigan K, Tooher R (2001) Relationships among speech perception, production, language, hearing loss, and age in children with impaired hearing. J Speech Lang Hear Res 44:264–285.

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

Bogen JE Bogen GM (1976) Wernicke's region — where is it? Ann NY Acad Sci 280:834–843.

Brungart DS, Simpson BD, Ericson MA, Scott KR (2001) Informational and energetic masking effects in the perception of multiple simultaneous talkers. J Acoust Soc Am 110:2527–2538.

Calvert GA, Bullmore ET, Brammer MJ, Campbell R, Williams SC, McGuire PK, Woodruff PW, Iversen SD, David AS (1997) Activation of auditory cortex during silent lipreading. Science 276:593–596.

Chan D, Fox NC, Scahill RI, Crum WR, Whitwell JL, Leschziner G, Rossor AM, Stevens JM, Cipolotti L, Rossor MN (2001) Patterns of temporal lobe atrophy in semantic dementia and Alzheimer's disease. Ann Neurol 49:433–442.

Crinion J, Blank S, Wise R (2003) Central neural systems for both narrative speech comprehension and propositional speech production.

Devlin JT, Russell RP, Davis MH, Price CJ, Wilson J, Moss HE, Matthews PM, Tyler LK (2000) Susceptibility-induced loss of signal: comparing PET and fMRI on a semantic task. Neuroimage 11:589–600.

Drullman R (1995) Temporal envelope and fine structure cues for speech intelligibility. J Acoust Soc Am 97:585–592.

Faulkner A, Rosen S, Wilkinson L (2001) Effects of the number of channels and speech-to-noise ratio on rate of connected discourse tracking through a simulated cochlear implant speech processor. Ear Hear 22:431–438.

Friederici AD, Ruschemeyer SA, Hahne A, Fiebach CJ (2003) The role of left inferior frontal and superior temporal cortex in sentence comprehension: localizing syntactic and semantic processes. Cereb Cortex 13:170–177.

Hall DA, Haggard MP, Akeroyd MA, Palmer AR, Summerfield AQ, Elliott MR, Gurney EM, Bowtell RW (1999) 'Sparse' temporal sampling in auditory fMRI. Hum Brain Mappp 7:213–223.

Hickok G Poeppel, D (2000) Towards a functional neuroanatomy of speech perception. Trends Cogn Sci 4:131–138.

Keidser G Grant, F (2001) Comparing loudness normalization (IHAFF) with speech intelligibility maximization (NAL-NL1) when implemented in a two-channel device. Ear Hear 22:501–515.

Kuest J Karbe H (2002) Cortical activation studies in aphasia. Curr Neurol Neurosci Rep 2:511–515.

Mazoyer B, Dehaene S, Tzourio N, Frak V, Cohen L, Murayama N, Levrier O, Salamon G, Mehler J (1993) The cortical representation of speech. J Cogn Neurosci 5:467–479.

Miller G (1951) Language and communication. New York: McGraw-Hill.

Mummery CJ, Ashburner J, Scott SK, Wise RJ (1999) Functional neuroimaging of speech perception in six normal and two aphasic subjects. J Acoust Soc Am 106:449–457.

Peters RW, Moore BC, Baer T (1998) Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people. J Acoust Soc Am 103:577–587.

Price CJ, Friston KJ (1997) Cognitive conjunction: a new approach to brain activation experiments. Neuroimage 5:261–270.

Remez RE, Rubin PE, Pisoni DB, Carrell TD (1981) Speech perception without traditional speech cues. Science 212:947–949.

Rosen S, Finn R, Faulkner A (2002) Plasticity in speech perception: spectrally-rotated speech, revisited. Association for Research in Laryngology midwinter meeting.

Sartori G Umilta C (2000) How to avoid the fallacies of cognitive subtraction in brain imaging. Brain Lang 74:191–212.

Schlosser MJ, Aoyagi N, Fulbright RK, Gore JC, McCarthy G (1998) Functional MRI studies of auditory comprehension. Hum Brain Mapp 6:1–13.

Scott SK Johnsrude IS (2003) The neuroanatomical and functional organization of speech perception. Trends Neurosci 26:100–107.

Scott SK, Wise R (2003a) The functional neuroanatomy of sublexical processing in speech perception. Cognition (in press).

Scott SK, Wise R (2003b) PET and fMRI studies of the neural basis of speech perception. Speech Commun (in press).

Scott SK, Blank CC, Rosen S, Wise RJ (2000) Identification of a pathway for intelligible speech in the left temporal lobe. Brain 123:2400–2406.

Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M (1995) Speech recognition with primarily temporal cues. Science 270:303–304.

Turner RS, Kenyon LC, Trojanowski JQ, Gonatas N, Grossman M (1996) Clinical, neuroimaging, and pathologic features of progressive nonfluent aphasia. Ann Neurol 39:166–173.

van Wijngaarden SJ, Steeneken HJ, Houtgast T (2002) Quantifying the intelligibility of speech in noise for non-native talkers. J Acoust Soc Am 112:3004–3013.

Vouloumanos A, Kiehl KA, Werker JF, Liddle PF (2001) Detection of sounds in the auditory stream: event-related fMRI evidence for differential activation to speech and nonspeech. J Cogn Neurosci 13:994–1005.

Wernicke C (1874) Der Aphasische Symptomenkomplex. Breslau: Cohn, Weigert.

Wise R, Chollet F, Hadar U, Friston K, Hoffner E, Frackowiak R (1991) Distribution of cortical neural networks involved in word comprehension and word retrieval. Brain 114:1803–1817.

Wise RJ, Scott SK, Blank SC, Mummery CJ, Murphy K, Warburton EA (2001) Separate neural subsystems within 'Wernicke's area'. Brain 124:83–95.

Wong D, Pisoni DB, Learn J, Gandour JT, Miyamoto RT, Hutchins GD (2002) PET imaging of differential cortical activation by monaural speech and nonspeech stimuli. Hear Res 166:9–23.

Zahn R, Huber W, Drews E, Erberich S, Krings T, Willmes K, Schwarz M (2000) Hemispheric lateralization at different levels of human auditory word processing: a functional magnetic resonance imaging study. Neurosci Lett 287:195–198.

Zatorre RJ, Evans AC, Meyer E, Gjedde A (1992) Lateralization of phonetic and pitch discrimination in speech processing. Science 256:846–849.